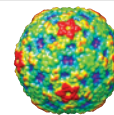


THIS WEEK

EDITORIALS

WORLD VIEW North–south divide could cripple European science **p.123**

CLIMATE Pacific salmon can stand the (moderate) heat **p.124**



DISEASE Respiratory-virus structure reveals drug pocket **p.124**

Reasons to be cheerful

As two new fronts in the war on disease demonstrate, creativity remains a key weapon for scientists in the hunt for drugs.

How many of the lists of ‘things to watch for in 2015’ dared to predict progress in the war on antibiotic resistance? After all, 2014 was the year when awareness of the antibiotic problem spilled onto the front pages, led by a gloomy landmark report by the World Health Organization released in April. In a vote to determine the focus of a new research effort funded by the UK government, the British public named antibiotic resistance as one of the most pressing challenges of modern times. The outlook for public health and the future was grim.

Well, whisper it, but there may finally be some good news. In a paper published on *Nature*’s website this week, a team of researchers announces the discovery of a brand new antibiotic (L. L. Ling *et al.* *Nature* <http://dx.doi.org/10.1038/nature14098>; 2015). Called (by them) teixobactin, the compound is produced by a specific type of soil bacterium — the cultivation of which was previously impossible (see also G. Wright *Nature* <http://dx.doi.org/10.1038/nature14193>; 2015). There could be more undiscovered antibiotics out there. There could be lots more.

Just as important as its discovery is the finding that teixobactin triggered no detectable genetic resistance in the bacteria it targets. With creditable understatement, the authors of the paper conclude: “The properties of this compound suggest a path towards developing antibiotics that are likely to avoid development of resistance.”

The discovery, of course, does not get us off the hook. Years of testing lie ahead and, even under the most optimistic scenario for teixobactin, growing resistance to other antibiotics remains a serious problem. The good news — and at this early stage, it is good news — must not drain

momentum from emerging efforts by policy-makers and others to tackle the dreadful and short-sighted squandering of precious antibiotic resources.

The positivity does not end there. In a *News* story on page 130, we detail the promising results of another drug that could help to address a crucial clinical need. This time, the drug — ketamine — is far from new. An anaesthetic used on the battlefields of the Vietnam War, it is now drawing attention for its potential to treat depression.

This potential is colossal, and much of that comes from the demand. As *Nature* highlighted in a special issue last year (see nature.com/depression), depression causes a greater burden than any other condition, yet the outlook at present is bleak. New drugs are proving difficult to come by, which is one reason that existing compounds such as ketamine are under the spotlight. Encouraging — if early — trial results suggest that the drug is both effective and fast-acting, even in people who have tried other therapies with little success.

Just as with teixobactin, much could still derail the early promise of ketamine for treating depression (and other mental-health conditions such as suicidal behaviour and bipolar disorder). But given the number of people who rely on antibiotics and who are desperate for help with depression, early promise is a lot better than the pessimistic messages that have previously dominated. So, with a smile and an optimistic attitude, here is a prediction for the new year: both ketamine and teixobactin could yet feature in the ‘highlights of 2015’ lists when they appear in December. Not bad for the second week of January. ■

Listen up

*Human echolocation kicks off the *Nature* podcast’s new series on sound science*

What is it like to be a bat? To sleep upside down, eat insects and use sound to ‘see’? The question is not as frivolous at it seems. Philosophers debate it at length as a way to think about minds and experiences. Still, a classic philosophy paper from 1974 argues that such an experience is ultimately off-limits for humans (T. Nagel *Philos. Rev.* **83**, 435–450; 1974). “It will not help to try to imagine,” it says, “that one has very poor vision, and perceives the surrounding world by a system of reflected high-frequency sound signals.”

Fiona Gameson does not have to imagine. Blind since the age of three, when she had both eyes removed because of a rare childhood cancer called retinoblastoma, she has learned to echolocate. She can navigate and see with sounds, just as bats do.

Fiona echolocates with a series of clicks — she pulls her tongue away

from the roof of her mouth to make a sharp ‘tick-tick’ sound. Then she listens for the echoes that bounce off objects around her, revealing their physical properties. In this way, she can navigate an unfamiliar house, for example, to find the bathroom at night without waking her hosts.

Words do not reflect her subtlety and skill. Only hearing her echolocate can do her story justice. Perhaps a more useful question for the rest of us, then, is what is it like to be Fiona Gameson? Showing us is the aim of the first episode in a new series of audio documentaries on *Nature*’s podcast that explore sound science in the medium that suits it best.

The series, *Audiofile*, begins next week. The inaugural episode explores Fiona’s remarkable ability, what we can really know about being a bat, and the limits of human perception. Subsequent monthly episodes will offer listeners sound-rich stories on the effects of noise on human health, archaeoacoustics and the impact of music on science. They will be available at go.nature.com/6stdre from 12 January.

Acoustic science is wide in scope and often practical in focus: at the twice-yearly meetings of the Acoustical Society of America, sessions exploring marine-mammal communication frequently run alongside lectures on architectural design and noise pollution. Until now, however, experiments in acoustics have been rarely showcased in *Nature*. These stories deserve to be heard. Audiophiles, lend us your ears. ■



Economic divide taking toll on European science

The Horizon 2020 programme threatens to siphon away the best scientists from southern Europe, argues Colin Macilwain.

Just before Christmas, European Union (EU) legislators delivered an unexpected bonus for European scientists: a generous 2015 budget for the flagship Horizon 2020 funding programme. The scheme, which runs from 2014 to 2020, seems to have wind in its sails, and has already proved immensely popular with researchers.

There is trouble ahead, however. The benefits of the programme are unevenly spread. Horizon 2020 will mainly finance countries and regions that are already doing well, but it will not do much for the other half of Europe, which has steadily weakening research and innovation capacities. This conflicts with a central mission of the EU Framework programmes, of which Horizon 2020 is the eighth version.

Apart from doing good research, the goal of the programmes has always been to develop closer, deeper ties across Europe's diverse research community. With the enlargement of the EU to the east, this became more difficult. Now, with economic mayhem in the south, this goal is in danger of receding out of sight.

Although Horizon 2020 only started last year, its main pillars were established in 2008, when most European countries were committed to increasing science spending. It concentrated on addressing 'grand challenges', such as climate change and healthy ageing, and on supporting 'research excellence'.

The scheme was well suited for a time of growth. But its architects could not know that by the time their grand plan was put into action, half of Europe would be falling off an economic cliff. Financial meltdown has drastically changed the topography of European research, and after decades of gentle convergence, the system is diverging sharply.

According to the European University Association's Public Funding Laboratory, for example, public money for universities has fallen by more than 10% in 10 European countries since 2008. In Italy it is down by 20%, and in Greece and Hungary by more than 40%. Funding has shot up by one-fifth, meanwhile, in Germany, Norway and Sweden, with Austria and Belgium not far behind.

The European Research Area (ERA) now functions very differently from when it was inaugurated in 2000. It was supposed to facilitate a single area within which students, researchers and ideas could freely circulate. Some aspects of this have been implemented: university recruitment, for example, is far more open and international than before.

However, this adjustment was not meant to instigate what we have now: a one-way flow of talent from the south and east to the north. If maintained for much longer, this haemorrhage of people and ideas can only entrench the divide in economic competitiveness that triggered the Eurocrisis in the first place.

In 2013, the most recent year for which

numbers are available, Germany and the United Kingdom were the top two recipients of Framework funding, receiving €1.1 billion (US\$1.35 billion) each. The Netherlands got €560 million, much the same as Italy, which has almost four times the population. Poland got €67 million; Romania just €17 million.

The EU has a mechanism, separate from Horizon 2020, that is designed specifically to address this sort of imbalance: 'structural funds' for infrastructure projects in poorer regions. These used to be aimed primarily at bringing eastern Europe up to speed with the west, but they are increasingly needed in the south, as well as in parts of the northern fringe.

For the 2014–20 period, around €50 billion in structural funds is meant to be made available for research and innovation-related projects, such as new laboratory buildings. But the use of structural funds for research purposes has not worked well in the past. I have toured glass palaces, constructed in improbable locales, with no researchers budgeted for and scant prospect of them ever arriving. Such projects are often backed by local political leaders who have weak understanding of what breeds innovation, wrongly believing that fancy buildings with Wi-Fi will do the job.

Commission officials know that €50 billion, working in concert with the total Horizon 2020 budget of €70 billion, could achieve a lot — in theory. But the practice is more troubling. Unlike Horizon 2020 projects, structural-fund projects are selected locally, with less EU oversight. Such infrastructure projects need political backing in

their own regions and take years. But even when they are finished, local researchers may be as unlikely as before to win fiercely competitive Horizon 2020 grants.

The EU has set up a small project, called Stairways to Excellence, to help applicants marry the two approaches. But what is needed is more fundamental reform: a thorough re-alignment that would use structural funds to build up research teams that are already almost good enough to partner in Horizon 2020 projects, together with a dash of 'positive discrimination' to favour these partners within Horizon 2020. This cannot happen under the current structure, and that stores up problems for the future.

Like other instruments of European integration, Horizon 2020 will not work if the centrifugal forces pulling the continent apart exceed the centripetal ones holding it together. EU research policy has always been a careful balance between 'excellence' and 'cohesion'. Right now, much more attention needs to be paid to the latter. ■

Colin Macilwain writes about science policy from Edinburgh, UK.
e-mail: cfmworldview@gmail.com

**THIS HAEMORRHAGE OF
PEOPLE AND
IDEAS
CAN ONLY ENTRENCH
THE DIVIDE
IN ECONOMIC
COMPETITIVENESS.**

➔ **NATURE.COM**
Discuss this article
online at:
go.nature.com/crgkk6

RESEARCH HIGHLIGHTS

Selections from the
scientific literature

ENGINEERING

Muscle sensor controls robot arm

A robot arm can mimic the movements of a human one, thanks to a device that measures muscle activity.

Muscle sensors are typically embedded in the tissue, but the flexible device created by Ifor Samuel and his colleagues at the University of St Andrews, UK, is worn on the operator's arm. An organic light-emitting diode in the sensor emits red light, which penetrates skin, bounces off muscle tissue and is detected by light-sensitive polymers in the device.

The sensor, unlike those that measure muscle electrical activity, can distinguish between types of muscle contraction, because light is scattered differently from muscle fibres when a person is lifting or pushing.

The team showed how the sensor could be used to control a robot arm, and hope that the device will one day be able to control prosthetic limbs using residual muscle.

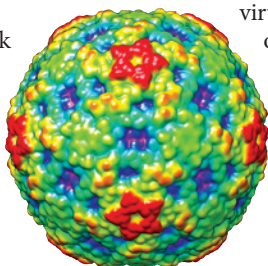
Adv. Mater. <http://doi.org/f2wpmh> (2014)

STRUCTURAL BIOLOGY

Cold drug fits in virus's pocket

An antiviral drug developed for the common cold could be effective against another virus that affected more than 1,000 US children in 2014.

An enterovirus called EV-D68 caused respiratory illness, sometimes severe, in an outbreak last August. Michael Rossmann and his colleagues at Purdue University in West Lafayette, Indiana, determined the



crystal structure of the virus (pictured) and found that, like other enteroviruses, it has a pocket in one of its proteins.

When the team worked out the structure of EV-D68 in conjunction with pleconaril, a drug that combats the related cold-causing rhinovirus, they found that the drug fitted into this pocket. In cultured cells infected with EV-D68, the

compound inhibited the virus better than two other antivirals, suggesting that pleconaril could be a candidate for treating this enterovirus.

Science 347, 71–74 (2015)

CLIMATE-CHANGE BIOLOGY

Salmon adapt to warmer waters

Pacific chinook salmon stand a good chance of adapting to higher temperatures in a scenario with modest warming, but could be wiped out under projections of maximum warming.

Bryan Neff at the University of Western Ontario in London, Canada, and his colleagues exposed the offspring of wild-caught chinook salmon (*Oncorhynchus tshawytscha*; pictured) to present-day temperatures and to those that are a few degrees higher, and measured the animals' maximum heart rate. They found that the fish

adapted their cardiac capacity, and were still able to reach maximum heart rate at temperatures 2°C warmer than today's. However, under conditions 4.4°C warmer than the present day, the salmon began to experience heart failure.

The authors predict that by the year 2100, chinook salmon populations will face a 17% chance of catastrophic loss under moderate warming, but a 98% chance of such loss if warming reaches a maximum.

Nature Clim. Change <http://doi.org/xws> (2014)



HUMAN EVOLUTION

Light bones linked to less activity

Low-density bones evolved only relatively recently in modern humans as they shifted from foraging to agriculture and became more sedentary, according to two studies.

Habiba Chirchir of the Smithsonian Institution in Washington DC, Brian Richmond of the American Museum of Natural History in New York and their colleagues analysed upper and lower limb bones from modern and early humans, several fossil ancestor species and chimpanzees. They found that bone density,

especially in lower limb joints, decreased in humans only in the past 12,000 years or so.

In a separate study, Timothy Ryan at Pennsylvania State University in University Park and Colin Shaw at the University of Cambridge, UK, examined bone structure in the hip joint in primates and in four human populations from several hundred to several thousand years ago. They discovered that bones from forager populations were much thicker and denser than those from agricultural groups, and were similar to those of wild non-human primates. The results support the idea that physical activity is important for bone strength.

MARK CONLIN/GETTY

YUE LIU & MICHAEL G. ROSSMANN, PURDUE UNIV.

at Johns Hopkins University in Baltimore, Maryland. The data have yielded new ways to classify tumours and pointed to previously unrecognized drug targets and carcinogens. But some researchers think that sequencing still has a lot to offer. In January, a statistical analysis of the mutation data for 21 cancers showed that sequencing still has the potential to find clinically useful mutations (M. S. Lawrence *et al.* *Nature* **505**, 495–501; 2014).

On 2 December, Staudt announced that once TCGA is completed, the NCI will continue to intensively sequence tumours in three cancers: ovarian, colorectal and lung adenocarcinoma. It then plans to evaluate the fruits of this extra effort before deciding whether to add back more cancers.

EXPANDED SCOPE

But this time around, the studies will be able to incorporate detailed clinical information about the patient's health, treatment history and response to therapies. Because researchers can now use paraffin-embedded samples, they can tap into data from past clinical trials, and study how mutations affect a patient's prognosis and response to treatment. Staudt says that the NCI will be announcing a call for proposals to sequence samples taken during clinical trials using the methods and analysis pipelines established by the TCGA.

The rest of the International Cancer Gene Consortium, slated to release early plans for a second wave of projects in February, will probably take a similar tack, says co-founder Tom Hudson, president of the Ontario Institute for Cancer Research in Toronto, Canada. A focus on finding sequences that make a tumour responsive to therapy has already been embraced by government funders in several countries eager to rein in health-care costs, he says. "Cancer therapies are very expensive. It's a priority for us to address which patients would respond to an expensive drug."

The NCI is also backing the creation of a repository for data not only from its own projects, but also from international efforts. This is intended to bring data access and analysis tools to a wider swathe of researchers, says Staudt. At present, the cancer genomics data constitute about 20 petabytes (10^{15} bytes), and are so large and unwieldy that only institutions with significant computing power can access them. Even then, it can take four months just to download them.

Stimulus funding cannot be counted on to fuel these plans, acknowledges Staudt. But cheaper sequencing and the ability to use biobanked biopsies should bring down the cost, he says. "Genomics is at the centre of much of what we do in cancer research," he says. "Now we can ask questions in a more directed way." ■



Marine biologist Sang-Mook Lee has pushed for academic involvement in South Korea's research ships.

OCEANOGRAPHY

Korea opens up its ocean science

Ships used mainly for seabed surveys will expand in focus.

BY MARK ZASTROW

South Korea's ocean-going research programme is changing tack. For more than two decades, it has focused on discovery and exploitation of minerals on the sea floor, but now a move is afoot to expand the research agenda. A 5,900-tonne ship — the *Isabu* — is being built with the capability to launch autonomous underwater vehicles, perform sea-floor-penetrating seismic surveys and collect sediment cores up to 40 metres long.

The current flagship, the 1,422-tonne *Onnuri*, spends about three-fifths of its time scouring the sea floor for mineral deposits under the direction of the deep-sea minerals group at the Korean Institute of Ocean Science and Technology (KIOST) in Ansan. That heavy economic emphasis is set by the Ministry of Oceans and Fisheries, which oversees KIOST as well as the nation's ports and shipping.

The ministry's hold is so complete that in 22 years of operation, no academic researcher outside KIOST has ever led a cruise. "This is really scandalous," says marine geophysicist Sang-Mook Lee of Seoul National University. Although scientists at his university and elsewhere have been able to work aboard the ship, they have been frustrated by a near-complete lack of say in where the *Onnuri*

goes or what research questions it pursues.

In March, that is set to change: KIOST will start to make *Onnuri*'s upcoming cruise tracks public, and will invite outside researchers to propose projects that can be done along the way, says Gi-Hoon Hong, who became the institute's president in August and has supported broadening the constituency for its research vessels. Eventually, time on the ships, which currently costs up to US\$12,000 per day, will be awarded through a merit-based system.

South Korea's focus on mineral exploration dates back to the founding of KIOST in the early 1970s, when the nation was in the middle of a decades-long economic boom. At the time, polymetallic nodules — balls of manganese and other metals such as iron, nickel and cobalt that accumulate on the sea floor — seemed a valuable potential resource. Although international interest in the minerals waned over subsequent decades, the South Korean government continued to fund research on the nodules and other sea-floor mineral deposits.

Securing marine mineral resources is "considered very important to the Korean people, because of the scarcity of land-based natural resources," says Jai-Woon Moon, the head of KIOST's deep-sea mineral research group. And rising prices for metals have renewed the world's interest: Nautilus Minerals of ▶

► Toronto, Canada, claims to be two years away from starting to extract gold, copper and other metals from a sea-floor site off Papua New Guinea.

But there is widespread scepticism of seabed mining. Lee calls the promised economic benefits “a big lie that the governments of Korea, China and India tell to their people”. Kyungsik Choi, a marine sedimentologist at Seoul National University, dismisses Nautilus Minerals and other commercial operations as not economically viable, and says that they will provide “nothing more than demonstrations”. Many in the Korean marine-science community say that KIOST itself is divided, with staff scientists chafing against the mining agenda imposed by the oceans ministry.

PUBLIC DEBATE

The long feud between the ministry and its critics reached a flashpoint in 2008, with the proposal to build the *Isabu*. The Korea Development Institute, a think tank in Sejong that is charged with evaluating major government projects, sided with the critics, saying that the economic benefits of sea-floor mining were uncertain. It approved the ship's construction, but on the condition that the academic community have access to it. A panel of researchers headed by Lee later recommended that the ship be managed by a committee with representatives from government, academia and industry. But in 2013, the oceans ministry transferred management to KIOST in a closed process. At the time, it was deemed most cost-effective for KIOST to both operate the vessel and direct its research, says Hyuntae Kim, director of the ministry's development division.

Disappointed by the move, Lee turned to a public forum. As one of the nation's most celebrated scientists and a well-known advocate for the rights of people with disabilities (he was paralysed from the neck down in a 2006 motor-vehicle accident), he testified in the national legislature on 24 October, accusing the oceans ministry of cutting a secret deal with KIOST. By the end of his appearance, then-minister Ju-Young Lee agreed to open up *Isabu* to the academic community, clearing a path for the merit-review process.

With the prospect of leading a major oceanographic cruise now open to him, Choi says that he hopes to take *Isabu* to the Indian Ocean. He wants to help nations in southeast Asia such as Myanmar and Bangladesh, which have extensive low-lying coastal deltas, to better understand the threat posed by sea-level rise and tsunamis.

Owing to his disability, Lee can no longer sail on cruises. But he says that he takes satisfaction in knowing that he has been able to wield his fame for a positive result. “I felt very good that, yes, this is compensation for my injury,” he says. “I felt redeemed.” ■



SCOTT HOUSTON/STORMA/CORBIS

Special K, or ketamine, a favourite drug at raves, is being considered as a treatment for depression.

PHARMACEUTICALS

Rave drug tested against depression

Companies and clinicians turn to ketamine to treat mental-health disorder as pipeline of new drugs dries up.

BY SARA REARDON

Ketamine, a psychoactive ‘party drug’ better known as Special K, has pharmaceutical companies riding high. Used clinically as an anaesthetic in animals and humans, it has proved an extremely effective treatment for depression, bipolar disorder and suicidal behaviour.

It also works incredibly fast. Unlike conventional antidepressants, which generally take weeks to start working, ketamine lifts depression in as little as two hours. “It blew the doors off what we thought we knew about depression treatment,” says psychiatrist James Murrough at Mount Sinai Hospital in New York City.

Companies are racing to develop patentable forms of the drug, and researchers are battling to understand how it affects the brain. An increasing number of clinicians are prescribing ketamine off-label for their patients, even as some of their colleagues worry that too little is known about its long-term effects.

The excitement over ketamine shows how badly new depression drugs are needed, says Thomas Insel, director of the US National Institute of Mental Health (NIMH) in

Bethesda, Maryland. Many drug companies have closed their mental-health divisions in the past five years, and there have been no significant advances in medication for depression in decades.

Today's most common antidepressants target the brain's serotonin or noradrenaline pathways (some target both). Ketamine blocks the signalling molecule NMDA, a component of the glutamate pathway, which is involved in memory and cognition. Before ketamine was studied, no one even knew that the pathway was involved in depression, Murrough says.

In 2013, his group published the largest trial of off-label ketamine carried out so far, with 73 participants. The trial found that the drug reduced depression 24 hours after treatment in 64% of patients who had tried three or more other medications with unsuccessful results. A second group received the sedative midazolam; in that case, the reduction was 28% (J. W. Murrough *et al. Am. J. Psychiatry* **170**, 1134–1142; 2013). Murrough's group is now imaging the brains of patients receiving ketamine treatment to try to dissect just how the drug works.

Murrough says that long-term studies of the

Proc. Natl Acad. Sci. USA [http://doi.org/xwr \(2014\)](http://doi.org/xwr (2014)); [http://doi.org/xwq \(2014\)](http://doi.org/xwq (2014))

DEVELOPMENTAL BIOLOGY

Cells stop dividing to heal wounds

Researchers have worked out why cells that stop dividing often release inflammatory signals that are linked to various age-related disorders.

Cells enter this 'senescent' state as tissues age, but it is not clear why they secrete numerous molecules, some of which cause inflammation. Judith Campisi at the Buck Institute for Research on Aging in Novato, California, and her team developed a mouse model in which they could isolate and study senescent cells. They found that blood-vessel and skin cells enter a senescent state at the site of a skin injury, where they promote healing by secreting a growth factor.

This could explain why senescent cells evolved the ability to secrete molecules, the team says.

Dev. Cell 31, 722–733 (2014)

NEUROSCIENCE

Direct route from gut to brain

Cells in the intestine can directly communicate with nerves, suggesting a way in which food and gut bacteria might affect the brain.

Intestinal cells called enteroendocrine cells regulate feeding behaviour, affecting the brain indirectly by secreting hormones. But these cells also have a protrusion called a neuropod. To find out whether this makes direct, physical contact with nerves in the intestine, Diego Bohórquez and his colleagues at Duke University in Durham, North Carolina, delivered a rabies virus (which typically infects neurons) to the colon of a mouse. They found that the virus first infected the intestinal cells and then passed into the animal's nervous system.

The neuropod might carry chemical signals from gut bacteria to the brain, and could allow viruses to spread from the intestine to the nervous system, the researchers suggest.

J. Clin. Invest. [http://doi.org/x2x \(2015\)](http://doi.org/x2x (2015))

IMMUNOLOGY

Fat cells fight bacterial infection

When a pathogenic bacterium invades the skin, fat cells are enlisted to help to kill the microbe, finds a study in mice.

Richard Gallo at the University of California, San Diego, and his team studied fat cells beneath the skin of mice that had been infected with *Staphylococcus aureus*. They found that fat cells increased in size and number in the early stages of infection and near the infection site. They also observed an increased amount of an antimicrobial compound, cathelicidin, when precursor mouse and human fat cells that had not been exposed to the bacterium developed into mature cells. The microbe enhanced the production of this molecule, which in turn slowed its growth.

Mice that could not generate fat cells were more susceptible to *S. aureus* infection.

Science 347, 67–71 (2015)

STEM CELLS

Diabetes reversed after transplant

Therapies that block part of the immune response in diabetic mice can prevent rejection of transplanted cells made from stem cells.

Some people with severe type 1 diabetes receive transplants of insulin-producing cells from healthy donors, but must take drugs that suppress much of the immune response, making them susceptible to cancer and infection. Jeffrey Bluestone at the University of California, San Francisco, and his colleagues suppressed immune responses in a more targeted

SOCIAL SELECTION

Popular articles on social media

Word-processing war ignites

Scientists generally welcome scrutiny of their data, but question their approach to word processing and sparks fly. A study suggesting that Microsoft Word is superior to the typesetting program LaTeX when it comes to efficiently creating academic documents has generated uproar online. "Def going to be the most controversial paper of the decade," tweeted Ben Lillie, a particle physicist who now directs The Story Collider, a series of science-storytelling podcasts and live shows. Indeed, many disagreed with the paper's finding. Steven Gibbons, a geophysicist at NORSAR, a geosciences research foundation in Kjeller, Norway, tweeted: "As a reviewer of scientific papers, I can confirm that ones produced in LaTeX are 500% better than those in Word for correct labelling of equations, figures, tables and including all references in the right order."

PLoS ONE 9, e115069 (2014)



Based on data from altmetric.com. Altmetric is supported by Macmillan Science and Education, which owns Nature Publishing Group.

➔ **NATURE.COM**
For more on popular papers:
go.nature.com/wyxb8z

way by treating diabetic mice with molecules that block T-cell stimulation. This prevented the immune system from rejecting transplanted pancreatic cells made from human embryonic stem cells, allowing the transplants to develop into insulin-secreting cells and reversing diabetes.

Targeting specific groups of immune cells could help to avoid the side effects associated with current immune-suppressing drugs, the authors say.

Cell Stem Cell [http://doi.org/xwg \(2014\)](http://doi.org/xwg (2014))

POPULATION GENETICS

Small numbers led to lemur demise

The number of giant lemurs in Madagascar may have been low even before humans pushed them to extinction.

Living lemur populations are found only in Madagascar, with the heaviest species weighing at most 7 kilograms. Giant species weighed up to 160 kg and vanished from the island after humans arrived around 2,000 years ago. A



team led by George Perry at Pennsylvania State University in University Park sequenced mitochondrial genomes from five extinct giant lemur species, including multiple samples from two species, *Megaladapis edwardsi* and *Palaeopropithecus ingens* (pictured). The authors found that both species had low genetic diversity — a sign of small population size.

The authors suggest that low population numbers, rather than large body size, made giant lemurs especially susceptible to extinction.

J. Hum. Evol. [http://doi.org/xwh \(2014\)](http://doi.org/xwh (2014))

➔ **NATURE.COM**
For the latest research published by Nature visit:
www.nature.com/latestresearch

SEVEN DAYS

The news in brief

RESEARCH

Stem-cell source

Stem cells that researchers claimed to have created by exposing ordinary cells to stress were probably embryonic stem cells that had been introduced into lab samples, according to an investigation of the claims. The RIKEN Center for Developmental Biology in Kobe, Japan, where the original claims were made, commissioned the investigation by outside experts. In a report released on 26 December last year, investigators noted the improbability of accidental contamination, but said that they could not determine whether misconduct had occurred. See go.nature.com/bk4atr for more.

POLICY

Blood ban to end

The US Food and Drug Administration has announced plans to end a lifetime ban on blood donation by men who have had sex with men. The long-standing ban was intended to prevent the spread of HIV, but has been criticized

as discriminatory by gay-rights activists. The proposed policy, revealed on 23 December 2014, will allow blood donations from men who have sex with men if at least a year has passed since their last sexual contact. The United Kingdom adopted a similar policy when lifting an analogous ban in 2011. See go.nature.com/s8oryr for more.

Lab-safety head

The US Centers for Disease Control and Prevention (CDC) has started recruitment for its first chief of laboratory safety. In July 2014, the agency revealed a series of accidents, including the mishandling of anthrax and a dangerous variety of the H5N1 influenza virus (see go.nature.com/lzqpwmm). On 23 December, scientists discovered that a technician might have been exposed to live Ebola virus in a CDC lab in Atlanta, Georgia.

Monarch status

The US Fish and Wildlife Service announced on 29 December last year that it will review whether to list a subspecies of the monarch butterfly (*Danaus plexippus*) as endangered



or threatened under the Endangered Species Act. Every year, many monarch populations migrate vast distances across North America (pictured), but population sizes have plummeted in recent years (see *Nature* **506**, 10; 2014). Habitat loss threatens the butterfly, and the use of agricultural pesticides has diminished the availability of milkweed, the caterpillar's only food source.

Bioethics concerns

In a letter on 22 December 2014, four members of the US Congress asked the National Institutes of Health (NIH) to review the bioethics of studies on monkeys at an NIH

laboratory. Animal-rights activists have campaigned against the experiments, which include studies that measure how separating young monkeys from their mothers affects alcohol consumption and responses to stress in the offspring later in life. The lawmakers have requested a bioethics report from the NIH by 27 February.

EVENTS

Balloon hopes burst

On 30 December last year, NASA brought down a scientific-balloon mission early because of a leak. The balloon had launched on 28 December from Antarctica and boasted a new 'super-pressure' design that was expected to help keep it aloft for 100 days or more. The 532,000-cubic-metre balloon was carrying a γ -ray telescope, known as the Compton Spectrometer and Imager, to hunt for high-energy photons streaming from the cosmos (see *Nature* <http://doi.org/x2q>; 2014). Two other balloon missions, ANITA and SPIDER, have launched successfully from Antarctica and are collecting data as planned.

BUSINESS

Biotech boom

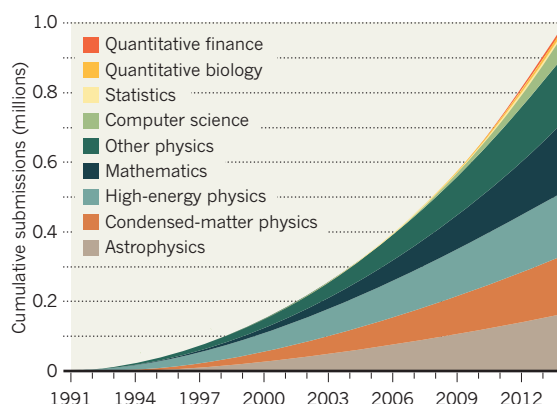
The cancer-treatment firm Juno Therapeutics launched the biggest initial public offering of 2014 for a biotechnology company, and finished the year worth more than US\$4.7 billion. Based in Seattle, Washington, the firm is developing ways to engineer immune cells to fight cancer (see *Nature* **516**, 156; 2014). Juno's public offering on 19 December was worth \$264.6 million, but the company's share price had climbed from \$24 to \$52.22 by the end of the year.

TREND WATCH

The preprint server arXiv.org, to which physicists, mathematicians and computer scientists routinely upload manuscripts to share their findings publicly before peer review, surpassed 1 million research articles on 29 December 2014, after administrators processed manuscripts submitted over the Christmas period. Founded in 1991 by physicist Paul Ginsparg at Cornell University in Ithaca, New York, the repository now receives about 8,000 submissions each month. See go.nature.com/di3rwb for more.

GROWTH MEDIUM

To cope with projected increases in manuscript submissions, arXiv plans to add one digit to its paper identification numbers.

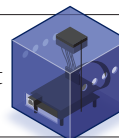


NEWS IN FOCUS

MEDICINE Triumph at completion of cancer genome 'atlas' **p.128**

OCEANOGRAPHY Science ahead for mineral-seeking Korean ship **p.129**

MENTAL HEALTH Party drug considered as treatment for depression **p.130**



PERSONAL SENSORS Fitness trackers are out; pollution trackers are in **p.136**

NIKOLAS GEORGIU/DEMOTIX/COREIS



University students in Athens protest against the government's plan for education and research reform.

POLITICS

Greek science haunted by hydra of problems

Leading researchers hang on despite austerity, but their Herculean efforts may not be enough.

BY ALISON ABBOTT

For chemical engineer Athanasios Konstandopoulos, it is as if all the myths of ancient Greece have come to life at once. The task of keeping up top-performing Greek labs such as his Aerosol and Particle Technology Laboratory (APTL) in Thessaloniki requires the strength of Hercules, he says, as well as the dogged persistence of Sisyphus, who was condemned for eternity to repeatedly roll a boulder up a hill and watch it roll down again.

Mortal power has so far maintained the scientific output of the APTL and other elite research centres in Greece, despite austerity

measures imposed in the wake of the nation's debt crisis in 2010. But five years on, with prolonged austerity pushing Greece into yet another political crisis, scientists are wondering how long that output can be kept up.

In 2014, budgets for research centres and universities in Greece were just one-quarter of their 2009 levels, and take-home salaries had been sliced by around one-third. This year begins with yet more cuts — even as the country implements a long-awaited law meant to reform the research landscape and make it more competitive. Qualified young professionals are leaving the country in unprecedented droves and researchers complain of being stifled by growing bureaucracy.

"The stress is building up for us," says Nektarios Tavernarakis, director of the FORTH Institute of Molecular Biology and Biotechnology (IMBB) in Heraklion, Crete, which churns out high-impact papers.

Although the 'troika' of organizations behind the austerity measures — the European Commission, the European Central Bank and the International Monetary Fund — considers science fundamental to economic recovery, it did not exclude science from the austerity measures. However, it insisted that cuts be coupled to reforms aimed at rejuvenating the country's generally lacklustre research and university systems.

In November, the law reforming research ►

► was passed, but it ended up being more conservative than researchers had hoped.

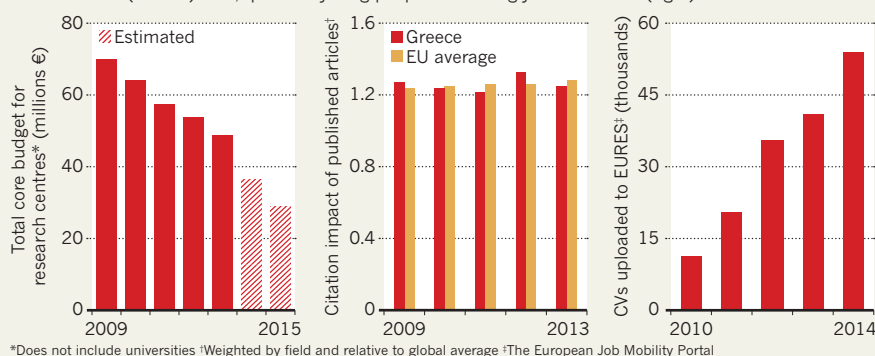
Its principle focus is on making it easier for scientists at universities and research centres to share facilities and to collaborate with industry, by removing bureaucratic obstacles. But researchers complain, for example, that the law did not create an independent grant agency to provide a source of regular support for basic research, akin to the US National Science Foundation or the UK research councils. Instead, Greek researchers depend almost entirely on external funders, such as the European Union, for grants.

Researchers also complain that they are subject to general public-sector rules for controlling public expenditure, even when that makes little sense. In one case that involved the centralization of payments within any public organization, the rules would have ended up costing the government money in grants lost. An appeal against those rules was successful. But the government then piled on extra bureaucracy: from January, researchers must report every expense higher than €1 (US\$1.2) into a centralized online system. “The rules come at you like the Hydra,” says Konstandopoulos. “You expend energy cutting off one of those heads, then another one grows back in its place.”

Theodore Fortsakis, rector of the University of Athens, the country’s largest university, says that if his budget is not increased, he will have to close some departments or centres in the second semester, which begins next month.

HEALTH OF GREEK SCIENCE IN NUMBERS

Government funding for research has plummeted (left) but scientists have maintained the quality of their research (middle). Still, qualified young people increasingly seek to leave (right).



The University of Crete is faced with a 2015 budget that is barely twice the institution’s 2014 electricity bill. Despite this, it came 48th in the *Times Higher Education* 2014 rankings of the world’s top 100 universities under 50 years old, and 5th for natural science in Europe in 2014 according to the CWTS Leiden Ranking of universities’ scientific performance.

Researchers say that the poor conditions are making it ever harder to recruit talented young scientists, even when positions arise. Economist Lois Labrianidis of the University of Macedonia in Thessaloniki has quantified the alarming rate at which qualified professionals are leaving the country. He calculates that 150,000 Greek professionals, including scientists, physicians and engineers — more than 50% of whom have a

PhD — now work in the rest of Europe and the United States. He says that the most qualified will not come back, and tens of thousands of others are actively seeking to leave (see ‘Health of Greek science in numbers’).

Not everyone has lost faith. Neurobiologist Marios Chatzigeorgiou last year accepted a group-leader position that will move him from the luxury of the MRC Laboratory of Molecular Biology in Cambridge, UK, to the IMBB. The uncertainty, he admits, is very worrying. But he believes that the multidisciplinary environment in Crete — which has a concentration of research institutes with diverse focuses ranging from marine science to computing — will be exciting. There, the spirit of Athena, goddess of wisdom, is alive and well. ■

GENETICS

End of cancer atlas prompts rethink

Geneticists debate whether focus should shift from sequencing genomes to analysing function.

BY HEIDI LEDFORD

A mammoth US effort to genetically profile 10,000 tumours has officially come to an end. Started in 2006 as a US\$100-million pilot, The Cancer Genome Atlas (TCGA) is now the biggest component of the International Cancer Genome Consortium, a collaboration of scientists from 16 nations that has discovered nearly 10 million cancer-related mutations.

The question is what to do next. Some researchers want to continue the focus on sequencing; others would rather expand their

work to explore how the mutations that have been identified influence the development and progression of cancer.

“TCGA should be completed and declared a victory,” says Bruce Stillman, president of Cold Spring Harbor Laboratory in New York. “There will always be new mutations found that are associated with a particular cancer. The question is: what is the cost–benefit ratio?”

Stillman was an early advocate for the project, even as some researchers feared that it would drain funds away from individual grants. Initially a three-year project, it was extended for five more years. In 2009, it received an

additional \$100 million from the US National Institutes of Health plus \$175 million from stimulus funding that was intended to spur the US economy during the global economic recession.

The project initially struggled. At the time, the sequencing technology worked only on fresh tissue that had been frozen rapidly. Yet most clinical biopsies are fixed in paraffin and stained for examination by pathologists. Finding and paying for fresh tissue samples became the programme’s largest expense, says Louis Staudt, director of the Office for Cancer Genomics at the National Cancer Institute (NCI) in Bethesda, Maryland.

Also a problem was the complexity of the data. Although a few ‘drivers’ stood out as likely contributors to the development of cancer, most of the mutations formed a bewildering hodgepodge of genetic oddities, with little commonality between tumours. Tests of drugs that targeted the drivers soon revealed another problem: cancers are often quick to become resistant, typically by activating different genes to bypass whatever cellular process is blocked by the treatment.

Despite those difficulties, nearly every aspect of cancer research has benefited from TCGA, says Bert Vogelstein, a cancer geneticist

at Johns Hopkins University in Baltimore, Maryland. The data have yielded new ways to classify tumours and pointed to previously unrecognized drug targets and carcinogens. But some researchers think that sequencing still has a lot to offer. In January, a statistical analysis of the mutation data for 21 cancers showed that sequencing still has the potential to find clinically useful mutations (M. S. Lawrence *et al.* *Nature* **505**, 495–501; 2014).

On 2 December, Staudt announced that once TCGA is completed, the NCI will continue to intensively sequence tumours in three cancers: ovarian, colorectal and lung adenocarcinoma. It then plans to evaluate the fruits of this extra effort before deciding whether to add back more cancers.

EXPANDED SCOPE

But this time around, the studies will be able to incorporate detailed clinical information about the patient's health, treatment history and response to therapies. Because researchers can now use paraffin-embedded samples, they can tap into data from past clinical trials, and study how mutations affect a patient's prognosis and response to treatment. Staudt says that the NCI will be announcing a call for proposals to sequence samples taken during clinical trials using the methods and analysis pipelines established by the TCGA.

The rest of the International Cancer Gene Consortium, slated to release early plans for a second wave of projects in February, will probably take a similar tack, says co-founder Tom Hudson, president of the Ontario Institute for Cancer Research in Toronto, Canada. A focus on finding sequences that make a tumour responsive to therapy has already been embraced by government funders in several countries eager to rein in health-care costs, he says. "Cancer therapies are very expensive. It's a priority for us to address which patients would respond to an expensive drug."

The NCI is also backing the creation of a repository for data not only from its own projects, but also from international efforts. This is intended to bring data access and analysis tools to a wider swathe of researchers, says Staudt. At present, the cancer genomics data constitute about 20 petabytes (10^{15} bytes), and are so large and unwieldy that only institutions with significant computing power can access them. Even then, it can take four months just to download them.

Stimulus funding cannot be counted on to fuel these plans, acknowledges Staudt. But cheaper sequencing and the ability to use biobanked biopsies should bring down the cost, he says. "Genomics is at the centre of much of what we do in cancer research," he says. "Now we can ask questions in a more directed way." ■



Marine biologist Sang-Mook Lee has pushed for academic involvement in South Korea's research ships.

OCEANOGRAPHY

Korea opens up its ocean science

Ships used mainly for seabed surveys will expand in focus.

BY MARK ZASTROW

South Korea's ocean-going research programme is changing tack. For more than two decades, it has focused on discovery and exploitation of minerals on the sea floor, but now a move is afoot to expand the research agenda. A 5,900-tonne ship — the *Isabu* — is being built with the capability to launch autonomous underwater vehicles, perform sea-floor-penetrating seismic surveys and collect sediment cores up to 40 metres long.

The current flagship, the 1,422-tonne *Onnuri*, spends about three-fifths of its time scouring the sea floor for mineral deposits under the direction of the deep-sea minerals group at the Korean Institute of Ocean Science and Technology (KIOST) in Ansan. That heavy economic emphasis is set by the Ministry of Oceans and Fisheries, which oversees KIOST as well as the nation's ports and shipping.

The ministry's hold is so complete that in 22 years of operation, no academic researcher outside KIOST has ever led a cruise. "This is really scandalous," says marine geophysicist Sang-Mook Lee of Seoul National University. Although scientists at his university and elsewhere have been able to work aboard the ship, they have been frustrated by a near-complete lack of say in where the *Onnuri*

goes or what research questions it pursues.

In March, that is set to change: KIOST will start to make *Onnuri*'s upcoming cruise tracks public, and will invite outside researchers to propose projects that can be done along the way, says Gi-Hoon Hong, who became the institute's president in August and has supported broadening the constituency for its research vessels. Eventually, time on the ships, which currently costs up to US\$12,000 per day, will be awarded through a merit-based system.

South Korea's focus on mineral exploration dates back to the founding of KIOST in the early 1970s, when the nation was in the middle of a decades-long economic boom. At the time, polymetallic nodules — balls of manganese and other metals such as iron, nickel and cobalt that accumulate on the sea floor — seemed a valuable potential resource. Although international interest in the minerals waned over subsequent decades, the South Korean government continued to fund research on the nodules and other sea-floor mineral deposits.

Securing marine mineral resources is "considered very important to the Korean people, because of the scarcity of land-based natural resources," says Jai-Woon Moon, the head of KIOST's deep-sea mineral research group. And rising prices for metals have renewed the world's interest: Nautilus Minerals of ▶

► Toronto, Canada, claims to be two years away from starting to extract gold, copper and other metals from a sea-floor site off Papua New Guinea.

But there is widespread scepticism of seabed mining. Lee calls the promised economic benefits “a big lie that the governments of Korea, China and India tell to their people”. Kyungsik Choi, a marine sedimentologist at Seoul National University, dismisses Nautilus Minerals and other commercial operations as not economically viable, and says that they will provide “nothing more than demonstrations”. Many in the Korean marine-science community say that KIOST itself is divided, with staff scientists chafing against the mining agenda imposed by the oceans ministry.

PUBLIC DEBATE

The long feud between the ministry and its critics reached a flashpoint in 2008, with the proposal to build the *Isabu*. The Korea Development Institute, a think tank in Sejong that is charged with evaluating major government projects, sided with the critics, saying that the economic benefits of sea-floor mining were uncertain. It approved the ship's construction, but on the condition that the academic community have access to it. A panel of researchers headed by Lee later recommended that the ship be managed by a committee with representatives from government, academia and industry. But in 2013, the oceans ministry transferred management to KIOST in a closed process. At the time, it was deemed most cost-effective for KIOST to both operate the vessel and direct its research, says Hyuntae Kim, director of the ministry's development division.

Disappointed by the move, Lee turned to a public forum. As one of the nation's most celebrated scientists and a well-known advocate for the rights of people with disabilities (he was paralysed from the neck down in a 2006 motor-vehicle accident), he testified in the national legislature on 24 October, accusing the oceans ministry of cutting a secret deal with KIOST. By the end of his appearance, then-minister Ju-Young Lee agreed to open up *Isabu* to the academic community, clearing a path for the merit-review process.

With the prospect of leading a major oceanographic cruise now open to him, Choi says that he hopes to take *Isabu* to the Indian Ocean. He wants to help nations in southeast Asia such as Myanmar and Bangladesh, which have extensive low-lying coastal deltas, to better understand the threat posed by sea-level rise and tsunamis.

Owing to his disability, Lee can no longer sail on cruises. But he says that he takes satisfaction in knowing that he has been able to wield his fame for a positive result. “I felt very good that, yes, this is compensation for my injury,” he says. “I felt redeemed.” ■



SCOTT HOUSTON/STORMA/CORBIS

Special K, or ketamine, a favourite drug at raves, is being considered as a treatment for depression.

PHARMACEUTICALS

Rave drug tested against depression

Companies and clinicians turn to ketamine to treat mental-health disorder as pipeline of new drugs dries up.

BY SARA REARDON

Ketamine, a psychoactive ‘party drug’ better known as Special K, has pharmaceutical companies riding high. Used clinically as an anaesthetic in animals and humans, it has proved an extremely effective treatment for depression, bipolar disorder and suicidal behaviour.

It also works incredibly fast. Unlike conventional antidepressants, which generally take weeks to start working, ketamine lifts depression in as little as two hours. “It blew the doors off what we thought we knew about depression treatment,” says psychiatrist James Murrough at Mount Sinai Hospital in New York City.

Companies are racing to develop patentable forms of the drug, and researchers are battling to understand how it affects the brain. An increasing number of clinicians are prescribing ketamine off-label for their patients, even as some of their colleagues worry that too little is known about its long-term effects.

The excitement over ketamine shows how badly new depression drugs are needed, says Thomas Insel, director of the US National Institute of Mental Health (NIMH) in

Bethesda, Maryland. Many drug companies have closed their mental-health divisions in the past five years, and there have been no significant advances in medication for depression in decades.

Today's most common antidepressants target the brain's serotonin or noradrenaline pathways (some target both). Ketamine blocks the signalling molecule NMDA, a component of the glutamate pathway, which is involved in memory and cognition. Before ketamine was studied, no one even knew that the pathway was involved in depression, Murrough says.

In 2013, his group published the largest trial of off-label ketamine carried out so far, with 73 participants. The trial found that the drug reduced depression 24 hours after treatment in 64% of patients who had tried three or more other medications with unsuccessful results. A second group received the sedative midazolam; in that case, the reduction was 28% (J. W. Murrough *et al. Am. J. Psychiatry* **170**, 1134–1142; 2013). Murrough's group is now imaging the brains of patients receiving ketamine treatment to try to dissect just how the drug works.

Murrough says that long-term studies of the

drug's effects should also be done before its use becomes widespread. And bioethicist Dominic Sisti of the University of Pennsylvania in Philadelphia worries that too many physicians already consider it a standard part of their armamentarium. The way in which ketamine should be administered still needs to be worked out, says psychiatrist Kyle Lapidus at Mount Sinai Hospital. He already prescribes ketamine off-label for some patients, and guesses that dozens of physicians across the country do the same. At therapeutic doses, it often produces a dissociative, out-of-body sensation that lasts less than an hour. At higher doses, recreational users report experiencing a 'K-hole', a deeply disoriented state accompanied by vivid hallucinations.

Companies hope to profit by developing patentable variations on ketamine for treating depression. A nasal spray containing a structural variant called esketamine earned a coveted 'breakthrough therapy designation' from the US Food and Drug Administration in 2013. The designation allows its manufacturer, Johnson & Johnson in New Brunswick, New Jersey, to fast-track esketamine through the regulatory process. The company plans to release the results of a 200-person study early this year; its head neuroscience researcher, Hussein Manji, says that initial results "look very good".

Last month, a company called Naurex, based in Evanston, Illinois, released results from a 386-person trial showing that its own ketamine-like drug, GLYX-13, successfully treated depression in about half of patients, without hallucinatory side effects. Roche of

"It blew the doors off what we thought we knew about depression treatment."

Basel, Switzerland, is also expected to release results early this year from a 357-person trial of a drug called decogluturant, which targets the glutamate pathway.

It is unclear why ketamine's psychoactive effects are considered a drawback, Sisti says. He questions the ethics of making patients pay more for a patented, non-dissociative drug if unmodified ketamine works just as well.

Ketamine's fast action is particularly promising for suicide prevention, says Carlos Zarate of the NIMH. Instead of being committed to institutions for weeks of treatment, people who have just attempted suicide might be treated with ketamine and released in days or even hours. Zarate has found that ketamine seems specifically to affect the desire to attempt suicide, whether a person is clinically depressed or not (E. D. Ballard *et al. J. Psychiatr. Res.* **58**, 161–166; 2014). That observation suggests that suicidal behaviour might be distinct from depression.

Zarate is using ketamine to treat around 50 people with depression, some of whom have suicidal thoughts, to study these effects.

Early this year, his group will begin a multi-year study of people who have attempted suicide within the previous two weeks, imaging their brain activity and comparing them with people who attempted suicide more than a year previously and with people with depression who have never attempted suicide. Those who have recently attempted suicide will be enrolled in a clinical trial of ketamine; at the same time, Zarate hopes to learn more about what an actively suicidal brain looks like. ■

CORRECTION

The News Feature 'Keeping the lights on' (*Nature* **515**, 326–329; 2014) incorrectly stated that the Boston Biomedical Research Institute went bankrupt and closed in 2013. It closed its doors under financial duress, but did not go bankrupt. The story also states that funds for indirect costs cannot be used to support researchers who lose grants or have yet to win one. Because the money from indirect costs goes into a general fund, institutions may spend it any way they wish, but these expenses cannot be reimbursed as indirect costs.



THE INSURMOUNTABLE GULF

Twenty-four years after the conflict ended, scientists and veterans are still fighting for recognition of Gulf War illness.

BY SARA REARDON

The meeting last April was supposed to be a scientific review, but the scene looked more like boxers lining up for their turn at a punchbag. The target was Robert Jesse, who at the time was deputy undersecretary for health at the US Department of Veterans Affairs (VA).

Veterans, scientists and VA administrators were meeting in Washington DC to discuss Gulf War illness, a complex disorder that affects some 250,000 veterans of the 1990–91 military operations in the Gulf. After 24 years, the condition is still the subject of intense controversy in the United States and the United Kingdom.

“From the beginning, the VA has refused to

honestly face the problems that face veterans,” said Joel Graves, a Gulf War veteran who until last year had served on a committee advising the VA on research priorities related to the illness.

Graves and others contend that the agency has refused to recognize Gulf War illness as a unique physiological condition, maintaining instead that it is psychosomatic or the result of stress. The VA, they claim, has obstructed research into Gulf War illness, stacked scientific review panels with members who would favour a psychological explanation and defanged the research advisory committee (RAC) that Graves served on. As a result, critics contend, thousands of soldiers have found it difficult to get a diagnosis or related

health benefits. At the meeting, James Binns, an attorney and chair of the RAC, called the VA’s actions “morally and intellectually bankrupt”.

When Jesse finally had a chance to speak, he flatly denied that the VA has clung to a psychological origin for the illness: “We have said unequivocally we do not believe that.” Rather, he and other VA officials believe that the veterans’ health problems are too complex to classify. Jesse then pointed to the VA’s multimillion-dollar research programme as proof that the agency takes the illness seriously.

Iraqi forces set fire to hundreds of oil wells as they retreated from Kuwait.

ABBAS/MAGNUM

The rift between scientists, veterans and the VA is deep, but upheavals at the agency in the past year may create an opportunity for reconciliation. “I don’t think it’s a hopeless cause,” says Victor Kalasinsky, an environmental toxicologist and the VA’s programme manager for Gulf War illnesses, “but people need to listen to one another.”

In 1990 and 1991, the United States deployed some 700,000 military personnel to the Gulf to form a coalition with the United Kingdom, Saudi Arabia and several other countries to expel Iraqi forces from Kuwait. The seven-month campaign resulted in few coalition casualties. But soon after returning home, about 30% of US veterans began to get sick. Their illnesses were difficult for doctors to understand.

They shared a cluster of symptoms — including severe fatigue, chronic pain, gastrointestinal disorders and cognitive problems. But few individuals had all of the symptoms, and there were many proposed causes. The destruction of chemical-weapons repositories was a leading suspect. Troops also marched past burning oil wells, slept in tents doused with pesticides and received new vaccines and pills to protect them from diseases and biological and nerve agents.

BRAIN AND BODY

Government agencies such as the VA, the US Department of Defense and the UK Ministry of Defence were initially reluctant to conclude that something unique was happening — veterans experience health problems after every war — and many doctors dismissed it as a form of post-traumatic stress disorder (PTSD) or even a psychosomatic condition.

But it was clear early on that something else was at play, says Roberta White, a neuroscientist at Boston University in Massachusetts who would eventually become scientific director of the RAC. Several early studies showed the same constellation of symptoms in veterans who had been deployed to the Gulf — particularly those exposed to chemical agents and pesticides — but not in veterans of the same era who had served elsewhere¹.

Evidence emerged that exposure to organophosphate chemicals in pesticides and the nerve agent sarin, and to nerve-gas prophylactics, were the probable cause. For example, a set of genetic analyses² by epidemiologist Robert Haley of the University of Texas Southwestern in Dallas found that symptoms were more severe in veterans who had a less active version of an enzyme that breaks down organophosphates. Veterans with fatigue and pain were more likely than controls to have brain-scan signals suggestive of nerve-fibre damage in certain areas of the brain^{3,4}. And veterans experienced cognitive and movement problems similar to those in farmers exposed to high levels of pesticides⁵.

But as research progressed, experts began to

suspect that US and UK government agencies were favouring a psychological explanation. This is hardly unprecedented: people with complex conditions such as chronic fatigue syndrome and the pain disorder fibromyalgia, which have some overlap in symptoms with Gulf War illness, have fought for years for acceptance by the scientific and medical community with limited success.

Veterans who felt that they had been ignored or marginalized after serving their country found allies among elected officials. In 1997, a

“VETERANS BELIEVE THEY’RE NOT GETTING THE CARE THEY NEED. OUR JOB IS TO GET TO THE BOTTOM OF THIS.”

congressional report⁶ determined that the VA and Department of Defense were concentrating too much on psychological causes, and called their research programme “irreparably flawed”. Among other findings, it charged that the agencies had lost or hidden chemical-exposure data, biased research funding towards psychiatric conditions, and made research impossible by automatically diagnosing veterans with PTSD and ignoring chemical exposures.

The report prompted Congress to create the RAC to evaluate the VA’s Gulf War research programme and advise on how to improve it. Its members would be appointed by the VA but would perform their analyses independently. Binns, a former defence-department official, was appointed as its first director in 2002, accompanied by 11 scientists and veterans.

In 2008, the RAC produced a 454-page report⁷ establishing that Gulf War illness was a distinct disorder tied to chemical exposures, with little to no role for stress and psychological factors. “I think that public opinion changed about Gulf War illness being a physical disease versus a psychiatric one” after this report, says White. “That was a big sea change for the field.”

The VA says that the report changed its thinking, too. But it continued to call the condition “chronic multisymptom illness” or “undiagnosed illness”, incensing veterans and researchers. The VA has resisted giving it an official name because it is difficult to establish diagnostic criteria for a unique syndrome. Epidemiologists have proposed several definitions, but they tend to be either too narrow to account for the diverse symptoms or so broad as to be meaningless.

Although anyone who served in the Gulf War is eligible for health care paid for by the agency, it can be difficult for veterans trying to claim

disability benefits to prove that their disability is connected to the war. Of the 54,000 claims for Gulf War-related illnesses filed as of last March, nearly 80% were denied, although more than half of those denied received compensation for other service-connected conditions.

In Britain, which sent more than 53,000 personnel to the Gulf War, the Ministry of Defence agreed in 2005 to use ‘Gulf War syndrome’ as an umbrella term “to provide an element of closure for those who have sought some acknowledgment that their ill-health is connected to their Gulf service”, says a government website. Still, advocacy groups such as the National Gulf Veterans and Families Association in Hull have found that many Gulf War veterans have difficulties claiming disability benefits.

In the United States, failing to adopt a standard definition for Gulf War illness has also complicated research. It has made it difficult to compare veterans who have the condition with those who do not, says RAC member Beatrice Golomb, a neurobiologist at the University of California, San Diego. What is more, the VA’s database defines Gulf War veterans as anyone who was deployed to the Gulf after 1990, and most in that category have not had the same chemical exposures as those in the 1990–91 conflict. That dilutes the sample, says Golomb.

Critics say that the VA has also hobbled an important research tool. Roughly every ten years, the agency conducts a survey of 30,000 veterans from the Gulf War era, asking about their health, symptoms and treatments. But the questions have changed every decade, making it hard to accurately track changes in health over time. The most recent version, sent out in 2012, included many questions about psychological stress, but not key questions that critics say are necessary to diagnose Gulf War illness. The RAC did not see the questionnaire before it went out, and when the committee pointed out the survey’s flaws, the VA responded that it would be too expensive to change it.

INSTITUTIONAL SUPPORT

As well as doing its own research and funding other researchers, the VA has commissioned a series of reports on Gulf War veterans’ health from the US Institute of Medicine (IOM). But critics from the RAC and elsewhere have claimed that the VA stacked the IOM committees with members who favoured psychological explanations for the disorder, and that it tailored the questions that the committees were to study or the pool of studies they were to review.

For the first time in 2010, an IOM committee specifically looked for a link between Gulf War service and chronic multisymptom illness. Although the report⁸ determined that a link exists, it said that there was not enough evidence from the human studies it had assessed to finger specific toxic exposures as a cause.

A 2014 IOM report⁹ urged the VA ►

A TENSE RELATIONSHIP

For more than two decades, the US Department of Veterans Affairs (VA) has been in conflict with researchers and veterans over Gulf War illness.

1990



JANUARY 1991: A US-led coalition bombs Iraqi chemical-weapons plants, triggering nerve-agent alarms at coalition bases to the south.

1995



NOVEMBER 1997: Prompted by veterans' grievances, a congressional report finds flaws in the VA's handling of Gulf War-related health issues, and calls for the formation of an independent advisory committee.

2000

JANUARY 2002: The Research Advisory Committee (RAC) on Gulf War Veterans' Illnesses is appointed to assess research, identify gaps and approaches, and recommend on programme needs.

2005



NOVEMBER 2008: The RAC releases a comprehensive 454-page report establishing the case for Gulf War illness and calling for US\$60 million in new annual funding for research.

2010

MARCH 2013: A VA whistle-blower alleges misconduct and cover-up of data pertaining to Gulf War-related health problems. An internal investigation finds no wrongdoing but requires publication of more data.

JUNE 2013: The VA announces the replacement of RAC members and revisions to its charter.

2015

► to start calling the disorder Gulf War illness. Although still hazy on the disease's origins, it recommended that the VA adopt two sets of diagnostic criteria that it could choose between, depending on its research needs. One, from the US Centers for Disease Control and Prevention, is broad; the other, developed by former RAC member Lea Steele, an epidemiologist at Baylor University in Waco, Texas, has stricter criteria and excludes people with known psychiatric disorders.

Although Binns sees the report as a great foundation for defining the disorder, some argue that the approach is a sign of equivocation. Bradley Flohr, senior adviser for compensation service at the VA, says it shows that even experts at the IOM are unable to agree.

The major limitation that the IOM has dealt with is the quality of the studies available for review, most of which are not comprehensive or have weaknesses in their methodology, says Kenneth Shine, former IOM president and head of several of its committees. He is not confident that any biomarker or discovery will make a more precise definition possible. "We have to say that the longer it goes from the deployment, the less likely it is that there will be a firm definition," he says, because the veterans are ageing and acquiring more illnesses that muddy the picture.

Kalasinsky says the agency does plan to adopt a definition for research purposes, but not necessarily for medical claims. As *Nature* went to press, the VA was still considering how to respond to the IOM's recommendations, nearly nine months after the report's release.

ALL CHANGE

Interactions between the VA and its RAC had long been adversarial (see 'A tense relationship'). But things came to a head in 2013 when, at a congressional hearing, members of the RAC testified that they had "no confidence" in the VA's research programme. Although the VA inspector-general found no evidence of wrongdoing, an investigation by the congressional committee on veterans' affairs has backed the RAC's claims that the agency was misappropriating money for Gulf War research and stacking IOM committees in its favour.

In June that year, the VA announced that it would be replacing most of the RAC. Members are supposed to serve for 2 years, but many had served for 12, so the VA said it was simply upholding the rules. It also rewrote the RAC's charter, stripping the mandate to evaluate the effectiveness of the VA's research programme and limiting the committee to reviewing current research and providing advice.

Existing members of the RAC see the move as retaliation. They say that they welcome new expertise on the committee, but they worry that the VA will try to appoint people who push a psychosomatic explanation. E-mails obtained by *Nature* show Jesse proposing candidates for three scientist positions; two have expertise in

psychosomatic illness and stress. Binns objected vigorously, and Jesse withdrew the nominations.

The VA would not discuss the search for candidates, but Kalasinsky denies that the decision to replace members was retaliatory and says that the changes to the charter were "not as draconian" as the RAC members believe.

But things are changing at the VA, which has suffered high-profile problems in the past year. The agency was thrown into chaos after an unrelated scandal in which VA hospitals falsified records to hide how long veterans were waiting for care. VA secretary Eric Shinseki resigned in May, and many officials, including Jesse, have moved positions. As a result, goals such as adopting a case definition for Gulf War illness have been delayed, says Kalasinsky.

At the RAC's most recent meeting, in September, several of the soon-to-be-dismissed members eagerly anticipated an appearance by the new VA secretary, Robert McDonald — a visit that Shinseki had never made. When he arrived, McDonald assured researchers and veterans that the VA believed that the veterans' suffering was real.

In an interview with *Nature*, McDonald said that he was busily educating himself on the illness. "Veterans believe they're not getting the care they need," he says. "Our job is to get to the bottom of this." Although he stopped short of saying that the disorder was something distinct, he said that the IOM's publications "seem to say it is very real".

Binns and Kalasinsky take McDonald's visit as a hopeful sign for reconciliation. But both acknowledge that it will be difficult. "The RAC makes recommendations. If they expect us to implement them all, that's not being realistic," says Kalasinsky. Nor is it accurate or helpful, he says, to continue suggesting that the VA supports the idea that Gulf War illness is psychosomatic. He says he does believe that veterans sometimes hear this from VA doctors, but that can be corrected. "We simply have to do a better job getting the word out and improve on our education programmes. I think the secretary is very committed to that." ■

Sara Reardon writes for *Nature* from Washington DC.

1. Steele, L., Sastre, A., Gerkovich, M. M. & Cook, M. R. *Environ. Health Perspect.* **120**, 112–118 (2012).
2. Haley, R. W., Billecke, S. & La Du, B. N. *Toxicol. Appl. Pharmacol.* **157**, 227–233 (1999).
3. Rayhan, R. U. et al. *PLoS ONE* **8**, e63903 (2013).
4. Rayhan, R. U. et al. *PLoS ONE* **8**, e58493 (2013).
5. Mackenzie Ross, S. J. et al. *Neurotoxicol. Teratol.* **32**, 452–459 (2010).
6. Committee on Government Reform and Oversight *Gulf War Veterans' Illnesses* House Report 105-388 (US Government Publishing Office, 1997).
7. Research Advisory Committee on Gulf War Veterans' Illnesses *Gulf War Illness and the Health of Gulf War Veterans* (Department of Veterans Affairs, 2008).
8. Institute of Medicine *Gulf War and Health: Volume 8* (National Academies Press, 2010).
9. Institute of Medicine *Chronic Multisymptom Illness in Gulf War Veterans: Case Definitions Reexamined* (National Academies Press, 2014).

UNSCOM/GETTY

RON EDMONDS/AP

DENNIS COOK/AP



POLLUTION PATROL

Step aside, fitness trackers. The next wave of personal sensors is giving people the ability to monitor the air they breathe.

BY KAT AUSTEN

When winter descends on Ulaanbaatar, Mongolia, the air turns foul. Here in the world's coldest capital city, residents light open fires of coal or wood to heat their uninsulated houses. Soot fills the skies, and people don face masks to ward off smog so thick it can hide buildings a few hundred metres away. "White clothing becomes grey after a few hours," says Munkhmandakh Myagmar, executive director of the Press Institute of Mongolia.

The city is one of the most polluted in the world, according to the World Health Organization (WHO) — and estimates suggest that particulate air pollution causes one-tenth of the city's deaths. But information about the extent of that pollution is limited and hard to find. The WHO's online database for pollution has readings from just one year for Ulaanbaatar,

ALTAF QADRI/AP

Joshua Apte travels around New Delhi in an auto rickshaw to measure air-pollution levels on the city's busy roads.

showing concentrations of harmful particles called PM₁₀, which have a diameter of 10 micrometres or less, and PM_{2.5}, with a diameter of 2.5 micrometres or less.

Eager to fill the knowledge gap, journalists from the Press Institute are taking matters into their own hands. In collaboration with the Earth Journalism Network — an international group of environmental journalists — Myagmar and her colleagues distributed five devices, each about the size of a child's lunch box, around the city in July.

Called DustDuinos, the devices measure particulate-matter concentrations and quickly upload the data to a public website. Despite some initial problems with charging and connectivity, preliminary results from a sensor in the city's centre showed that concentrations of PM₁₀ often surged to at least twice the WHO's recommended limit.

The DustDuino and other pollution sensors, some of which can be built for as little as US\$50, and instructions for which are available online, are part of the next wave in the environmental movement (see 'Sensors for the people'). Across the globe, journalists, advocacy groups, hackers and others are starting to use low-cost monitoring devices to vastly expand the amount of data that are publicly available on forms of air pollution such as particulate matter and toxic gases. The devices are easy to deploy and can complement data from official networks, which rely on sophisticated but sparsely distributed sensors. The 'citizen-science' approach aims to provide high-resolution measurements of air pollution where people actually live. Work is also under way to develop wearable sensors to monitor personal exposure levels.

Built on the principle of openness, such do-it-yourself (DIY) efforts are part of a push to democratize air-quality monitoring so that it no longer remains solely in the domain of governments and academic researchers. But advocates of the approach still have to convince conventional pollution researchers, who worry about the quality and usability of data from cheap sensors operated by relatively untrained people.

Still, everyone agrees that more resources need to go into monitoring air pollution, which kills around 7 million people a year. "It's the largest, single most important, health risk in the world," says Joshua Apte, an environmental researcher at the University of Texas at Austin, who sees an emerging role for cheap, plentiful pollution gauges. "The fact that you can buy 50 low-cost sensors for the cost of one regulatory sensor is a tremendously powerful thing."

POLLUTION PIGEONS

The roots of the movement go back to 2006, when Beatriz da Costa, an artist at the University of California, Irvine, strapped a small bundle of sensors onto homing pigeons. Da Costa had worked with engineers to develop the instrument package, which measured carbon monoxide and nitrogen oxides and tracked the pigeons' movements using a Global Positioning System (GPS) receiver. She published the data from her project, called PigeonBlog, with the aim of disrupting the status quo and giving the public a role in gathering data on pollution.

Around the same time, sensors for other uses were starting to emerge. More than two-thirds of US adults now say they use technology to track their heart rate and other health data; gadgets for the home sense water and electricity usage by the minute, and cities are employing sensors to track everything from pedestrian traffic to leaky pipes.

Options for grass-roots pollution monitoring were scarce until a few years ago: conventional air-quality sensors are

expensive or require training, and the data they provide are often inaccessible or hard to work with and share. That began to change when 'makers' — tinkerers who work in backyard sheds and collective hackspaces — started soldering together circuit boards and sensors to take on the challenge.

In 2011, a group of hackers, makers and artists who called themselves the Sensemakers gathered at meet-ups in Amsterdam and New York. Governments were not monitoring pollution at the local level, where it affects people, they wrote on their blog. The Sensemakers launched an online call for a community-developed sensor that would measure air pollution. With more than \$144,000 raised through the crowd-funding platform Kickstarter, the Sensemakers developed the Air Quality Egg, a device to measure temperature, humidity, carbon dioxide and nitrogen dioxide. The sensor costs \$185 — less than one-tenth the price of a mid-range device.

Similar efforts were emerging in Spain at around the same time. Tomas Díez Ladera, director of Fab Lab Barcelona at the Institute for Advanced Architecture of Catalonia, had dreams of citizens being able to monitor the air they breathe

"IT'S THE LARGEST, SINGLE MOST IMPORTANT, HEALTH RISK IN THE WORLD."

in real time and stream the data so that others in their community could benefit. After spending a year looking at what technology was available, Díez Ladera decided that he and his team would have to realize this dream themselves. Their early prototypes grew into a package of sensors called the Smart Citizen Kit (SCK) — which measures everything the Air Quality Egg does, as well as light intensity and noise. The group launched a website to encourage DIY-ers to build or buy the device and monitor local pollution.

Because open-source sensors such as these can be made anywhere, it is difficult to track how many have been produced. But according to their inventors, at least 35 DustDuinos, some 2,500 Air Quality Eggs and 1,000 SCKs have been deployed. Many of the data produced by these devices are openly available through online platforms such as Xively.

These efforts have already captured the attention of city officials and citizen groups. In early 2014, officials in Amsterdam provided 100 citizens with SCKs and instructions on how to use them in their neighbourhoods. In May, a community effort coordinated by FutureEverything, an innovation lab in Manchester, UK, created a network of sensors around that city's centre. The aim is to test how such data can enhance urban living by showing, for example, how efforts to encourage bicycling affect air quality.

DATA DIVIDE

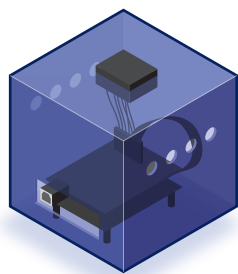
All this enthusiasm by proponents has not yet won over environmental researchers. The sensors currently used in the SCK and the Air Quality Egg are not up to the task of providing robust pollution data, says Ben Barratt, an air-quality scientist at King's College London, who helps to run the London Air Quality Network, an online resource for pollution data.

"Monitoring air-pollution levels is far more involved than the manufacturers and suppliers of cheap sensors suggest," Barratt says. The problem is that temperature, humidity and some gases skew the results from sensors such as those used in the Air Quality Egg and SCK, making it difficult to compare results between devices, he says.

DIY endeavours also lack input from pollution experts,

SENSORS FOR THE PEOPLE

How polluted is your home or neighbourhood? Until recently, it was difficult to answer that question because data were available only from networks of expensive sensors in relatively limited locations. The do-it-yourself movement has led to the emergence of low-cost sensors that can be purchased or built from online instructions.



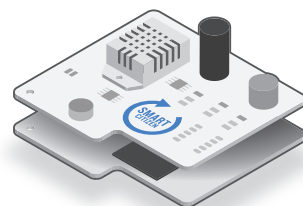
DUSTDUINO

Measures the concentration of pollution particles equal to or smaller than 10 micrometres (PM₁₀) and 2.5 micrometres (PM_{2.5}).



AIR QUALITY EGG

Measures nitrogen dioxide and carbon monoxide, humidity and temperature. Streams data online through a separate base station.



SMART CITIZEN KIT

Measures nitrogen dioxide and carbon monoxide, as well as light, humidity, temperature and noise pollution.

says Tim Chatterton, a pollution-policy researcher at the University of the West of England in Bristol, UK, who has worked with the UK government to monitor air quality. Professional technicians and scientists pay close attention to siting and maintaining sensors, he says. “Without due attention to these things, the data is essentially meaningless because it’s not comparable.”

These problems reduce the usefulness of such data in environmental advocacy, says Kirk Smith, an environmental health researcher at the University of California, Berkeley. The US Environmental Protection Agency (EPA) has strict rules about what holds up in court, he says, “and these monitors don’t meet their criteria — yet”.

The data are also often streamed and stored in ways that make them difficult to analyse and visualize to tell a coherent story, adds Smith, whose group has created particulate-matter sensors for use indoors. “One thing we learned in devel-

next release of the SCK, expected in early 2015, will be pre-calibrated and have better sensors, which should make its data reliable enough for comparison with official air-quality standards.

Apte’s previous work has shown just how important it is to understand how pollution levels vary at the human scale. Last year he spent four months driving through New Delhi’s frenetic roads on a rickshaw kitted out with a mid-range sensor called a DustTrak, which can deliver real-time feedback about pollution.

Apte wanted to understand how pollution levels change as people go about their daily business — and he discovered large variations. In traffic, for example, a passing truck spewing smoke can cause concentrations of PM_{2.5} to jump by as much as 50%.

During those rides, Apte found even greater fluctuation in the levels of harmful ultrafine particles, which are smaller than PM_{2.5}, and of pure carbon specks. “We’re interested in deploying networks of air-pollution monitors all around cities,” says Apte. “You couldn’t do this if you were paying \$10,000 per sensor.”

Other researchers are engaging with the citizen-sensing movement. Despite his concerns over the data quality, Barratt is advising projects such as the London Sustainability Exchange, which is working with communities to measure air pollution in the city.

The EPA is also starting to embrace the citizen-monitoring concept. Tim Watkins, acting deputy director of the EPA’s National Exposure Research Laboratory, wants to explore how cheaper, less accurate sensors can provide data that will complement the sparsely spaced, top-of-the-range kits. “This new tech is potentially very valuable. And it’s coming, whether or not we are investing or using it,” he says.

In 2013 the EPA announced the winners of a competition for developers to create low-cost wearable sensors that integrate air-quality measurements with health data. And in the next few months the agency will announce the winners of a \$4.5-million competition to fund research on community use of low-cost sensors to measure air quality.

As budget cuts cause governments to trim expensive sensor networks, citizen sensing will develop to fill the gap, says Schroyer. He foresees a day when people’s clothes will measure their exposure to carcinogens, their phones will sniff for polluting particles and drones will hover over cities, searching for natural-gas leaks. “Mobile, fabrics, health monitors — these are all quite possible,” he says. “There are hackers the world over working on these technologies right now.” ■

Kat Austen is a freelance writer in Berlin.

“THIS NEW TECH IS POTENTIALLY VERY VALUABLE. AND IT’S COMING, WHETHER OR NOT WE ARE INVESTING OR USING IT.”

oping our smart, cheap monitors is that they are producing a lot of data and it’s not clean. To get something useful out of the end takes a lot of work,” he says.

Given such concerns, the keepers of scientific data sets have not yet embraced the information produced by most DIY sensors. Barratt, for example, says he will not include data from citizen-sensing projects in the London Air Quality Network until the quality improves.

Matthew Schroyer, who developed the DustDuino, says that its data are comparable to those generated by more expensive sensors when averaged over a sufficient time span. An independent team of researchers reported that the sensor used in the DustDuino performs as well as a higher-cost sensor when sampling air quality over windows of 1 hour (D. M. Holstius *et al. Atmos. Meas. Tech. Discuss.* 7, 605–632, 2014). Schroyer, a technology developer and communications specialist in Champaign, Illinois, acknowledges that the DustDuino is not accurate enough for instantaneous data gathering, because its measurements contain too much noise.

The developers of the Air Quality Egg and the SCK also recognize that the sensors face teething problems, but they say that they are addressing them. Diez Ladera says that the

COMMENT

GEOLOGY From chemistry to conflict, the scramble for rare metals **p.142**



SOCIETY Why water has always been centre stage in China's politics **p.144**

PUBLISHING A call to pay peer reviewers for their time and effort **p.145**

POLLUTION Beijing's week of blue skies — good policy or bad business? **p.145**

US NATL LIB. MEDICINE



Max Cooper in Robert Good's lab at the University of Minnesota in the mid-1960s.

Fifty years of B lymphocytes

Alexander D. Gitlin and **Michel C. Nussenzweig** reflect on the discovery of two lineages of adaptive immune cells, and how it influenced vaccination, cancer therapy and the development of a class of antibody-based drugs.

When Max Dale Cooper joined Robert Good's laboratory at the University of Minnesota in Minneapolis in 1963, there were two camps in immunology. Neither was particularly fond of the other.

At the time, the central question in immunology was how vertebrates tailor their defences to bacteria and viruses, whose chemical structures show nearly unlimited diversity. Within two years of joining Good's laboratory, Cooper had made a discovery

about the cells that accomplish this task — lymphocytes — that proved essential to cracking the mystery, and ultimately to unifying the camps of his field.

Fifty years ago this week, Good, Cooper and their colleague Raymond Peterson published a paper¹ in this journal revealing that there are two types of lymphocyte. The insight shaped the course of modern immunology and influenced the study and care of immunodeficiency conditions, cancers of the immune system and the development of

monoclonal antibodies — powerful research tools and therapeutics.

CLONE WARFARE

In the 1960s, one camp of immunologists dealt mainly in chemical terms and had by then made considerable progress. This group had discovered that antibody molecules are proteins with two binding sites that recognize an extraordinary range of foreign molecules (antigens), even synthetic ones; and that antibodies are composed of two heavy and ►

► two light chains, with amino acids that vary at one end (the N terminus) but that are constant at the other end (the C terminus).

The second camp dealt with immunology at the cellular and whole-organism level. Here, the clonal-selection theory, which hypothesized that lymphocytes are diverse and that each is unique, or clonal, was becoming increasingly accepted. Each cell carries a unique surface receptor that, when bound by antigen, triggers the proliferation of that clone. This theory was developed in the late 1950s by Frank Macfarlane Burnet at the Walter and Eliza Hall Institute (WEHI) of Medical Research in Melbourne, Australia, and David Talmage at the University of Chicago in Illinois^{2,3}. It gave immunology a conceptual framework, yet proof of its existence and its workings remained elusive.

In 1961, Jacques Miller, also at the WEHI, demonstrated that a mouse with its thymus removed did not reject skin grafts from different strains of mice, as the animals normally would⁴. This suggested that the thymus was a crucial source of cells that mediate phenomena such as rejection, but his experiments were not universally accepted. At that time, antibodies were thought to derive only from lymphocytes (through plasma cells). But immunologists did not know whether and how the lymphocytes that produced antibodies and those involved in graft rejection were related.

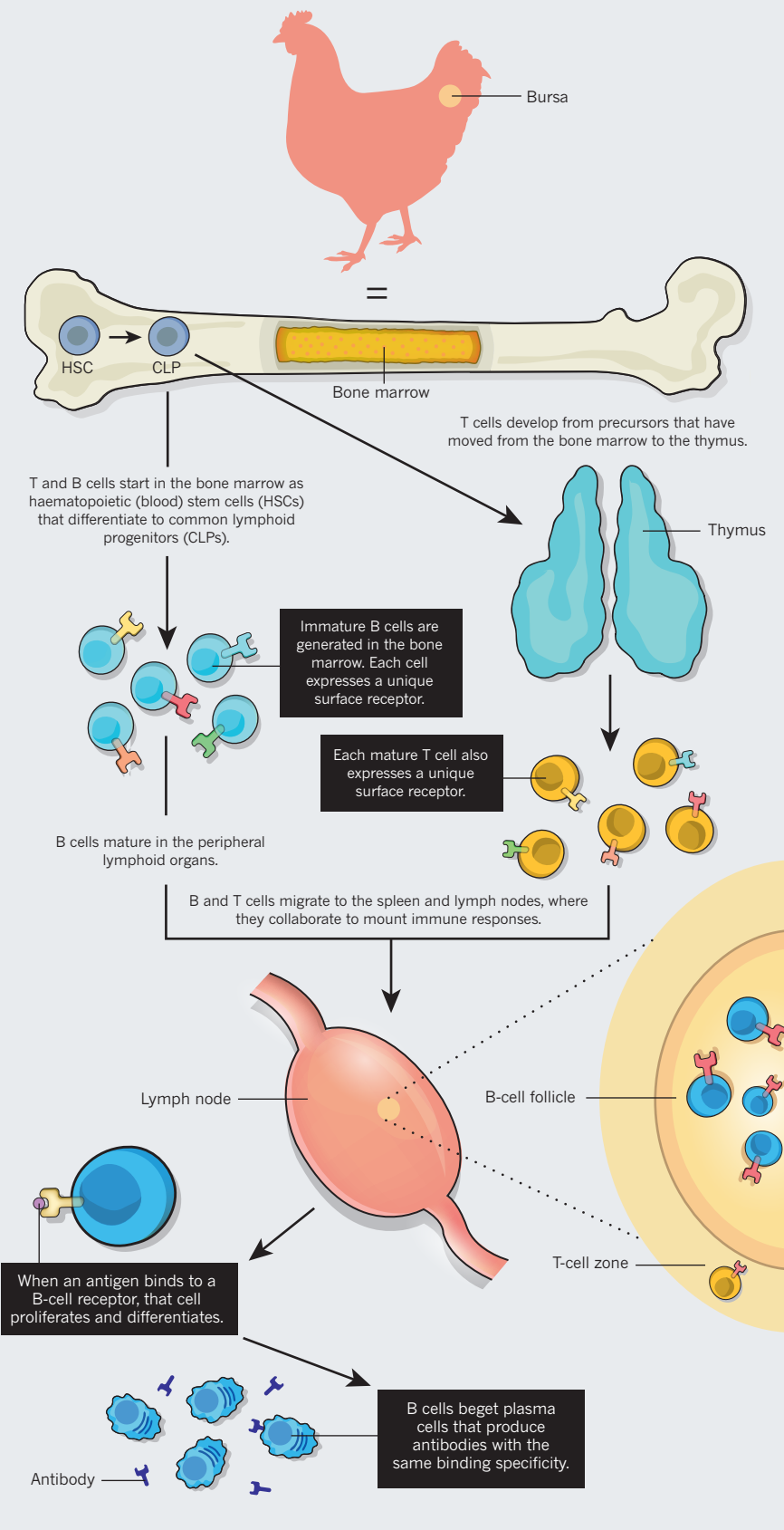
It was in this context that Cooper, a paediatrician and clinical immunologist, made a set of clinical observations that were an early clue to the existence of two lymphocyte lineages⁵. People with Wiskott–Aldrich syndrome, a disorder linked to the X chromosome, developed rampant herpes viral lesions associated with defective immunity. Yet these people had high levels of antibodies. By contrast, boys with the inherited immunodeficiency X-linked agammaglobulinaemia could control such viral infections even though they lacked antibody responses. This hinted that antibody-producing lymphocytes and thymus-derived lymphocytes that reject grafts (as in Miller's mouse experiments) might be two different things.

THE KEY EXPERIMENT

Cooper was one of the few immunologists who found clues to this paradox in a 1956 *Poultry Science* report⁶. It recounted an 'accidental' finding about the bursa of Fabricius, an organ unique to birds that at the time was thought to have a hormonal role in growth and development. The authors, Bruce Glick and his colleagues, wrote of their "suspicion regarding the importance of the bursa in antibody production". To study the role of the organ, Glick and his colleagues used chickens that had had their bursae removed for an unrelated experiment. To the researchers' surprise, these chickens showed

TWO LINES OF ATTACK

In mammals, the haematopoietic (or blood-forming) tissues serve the same immune function as the bursa in chickens.



remarkably poor antibody production.

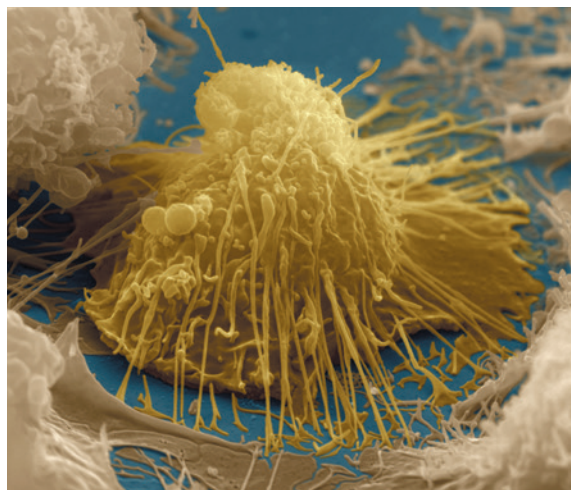
Good and Cooper decided to revisit the role of the bursa and the thymus in chickens. Up until that point, organ-removal studies in chickens had yielded conflicting results. Cooper reasoned that these differences could have arisen because some chickens might have developed immune cells before their organs were taken out⁵. His solution was to irradiate hatchlings a day after removing the organs, which eliminated any cells generated by a chick's thymus or bursa before hatching. These chicks would thus reveal the roles that these organs have in immune development.

The result, published in *Nature* on 9 January 1965, was spectacular¹. Irradiated chickens with no bursae produced no antibodies when injected with either bovine serum albumin, a protein derived from cattle, or the bacterium *Brucella abortus*. Serum from the birds completely lacked the major antibody classes. Yet the region of the white pulp in the spleen that is dependent on the thymus was intact. The dramatic phenotype was highly reminiscent of that of patients with X-linked agammaglobulinaemia.

The next year, Cooper and his colleagues extended their seminal study in an important paper published in the *Journal of Experimental Medicine*⁷. They used their irradiated chickens — without thymus and bursa — to elucidate the different functions of the immune cells produced by the two organs. They found that B (bursa-derived) cells are required for antibody responses, whereas T (thymus-derived) cells mediate delayed-type hypersensitivity reactions, graft-versus-host rejection and skin-graft rejection — all functions of cellular immunity.

The explanatory power of the two-lymphocyte model for immunodeficiency diseases was immense⁵. Because people with X-linked agammaglobulinaemia are deficient in antibody production but not in cellular immunity, their disease was probably due to a deficiency in B-cell development only. By contrast, the Swiss-type agammaglobulinaemia, in which both cellular and antibody-based immunity are severely compromised, was likely to be caused by a deficiency in a precursor cell common to both T and B lineages. Not surprisingly, clinicians were more receptive to Cooper's findings than were basic immunologists.

Next, the most important quest for Cooper was to find an organ in mammals equivalent to the bursa. Without it, the relevance and generality of his findings remained



A coloured scanning electron micrograph of a B lymphocyte.

controversial for nearly another decade. Scientists would often ask Cooper, "Where is the bursa equivalent this year, Max?"²⁵

Finding the equivalent organ to the bursa in mammals proved difficult. Initially, Cooper and his colleagues suspected that intestinal tissue was the source of B cells; they expended enormous effort on this dead end⁵.

The answer finally came in 1974, when Cooper, together with Martin Raff and John Owen at University College London, cultured mouse fetal liver cells from day 14 of gestation. After four to seven days in culture, B cells were generated⁸. At the same time, groups led by Gustav Nossal at the WEHI and Pierre Vassalli at the University of Geneva, Switzerland, made similar findings, using mouse bone marrow⁵. In other words, the haematopoietic, or blood-forming, tissues serve the same function in mammals as the bursa in chickens (see "Two lines of attack").

By the late 1970s, this understanding of the divergent lymphocyte lineages started to change the treatment of leukaemias and lymphomas. The cells of origin of these tumours could now be used to classify the cancers and to tailor therapies. Thus, acute lymphoblastic leukaemias could be B-cell- or T-cell-derived or neither; non-Hodgkin's and Burkitt's lymphomas are B-cell-derived. The effectiveness of various therapies for these cancers often dovetailed with their classification.

DRUGS AND VACCINES

Among the biggest impacts of B-cell biology was the invention of hybridomas — immortal antibody-producing cell lines. In 1975, Georges Köhler and César Milstein at the MRC Laboratory of Molecular Biology in Cambridge, UK, reported that they had fused B cells with a myeloma cell line⁹. The resulting hybrid cell continuously produced its specific antibody. This was the birth of monoclonal-antibody technology, garnering Köhler, Milstein and Niels Jerne the 1984

Nobel Prize in Physiology or Medicine.

The effect of this accomplishment is difficult to overstate. In the ensuing years, monoclonal antibodies permeated all of experimental biology. These molecules are potent and specific reagents that can be used to identify, isolate and perturb nearly any molecule or cell of interest. Clinically, monoclonal antibodies have become some of the most powerful diagnostics and therapeutics.

Cooper's discovery also set the stage for answering the fundamental question of clonal-selection theory: how is clonal diversity generated? The elegant molecular solution to this problem was elucidated¹⁰ by Susumu Tonegawa in 1976. B cells assemble a diverse set of antibodies by combining a set of three types of gene segment from a large pool of potential segments.

Finally, the antibodies for which so much had been elucidated at the chemical level were now understood at the genetic, cellular and organismal levels. By the mid-1980s, the clonal nature of the T-cell system had also been explained. Through their discovery of the B and T cell systems, Cooper and his colleagues set in motion a series of landmark findings that united the chemical and cellular camps of the discipline.

Key questions remain in B-cell biology. In particular, there are major unmet needs in vaccine development for HIV, influenza and many other infectious agents. Understanding how B cells are selected to differentiate into long-lived cells that provide protection from infection will guide us where empirical approaches have failed.

As these discoveries continue to affect human health, it remains important to remember their origins in an experiment performed 50 years ago, by Cooper, in chickens. ■

Alexander D. Gitlin is an MD–PhD student in the Weill Cornell/Rockefeller/Sloan-Kettering Tri-Institutional programme in New York City, New York, USA.

Michel C. Nussenzweig is a professor and HHMI investigator at Rockefeller University in New York, New York, USA.
e-mail: nussen@mail.rockefeller.edu

- Cooper, M. D., Peterson, R. D. A. & Good, R. A. *Nature* **205**, 143–146 (1965).
- Burnet, F. M. *The Clonal Selection Theory of Acquired Immunity* (Vanderbilt Univ. Press, 1959).
- Talmage, D. W. *Science* **129**, 1643–1648 (1959).
- Miller, J. F. A. P. *Immunol. Rev.* **185**, 7–14 (2002).
- Cooper, M. D. A. *Ann. Rev. Immunol.* **28**, 1–19 (2009).
- Glick, B., Chang, T. S. & Jaap, R. G. *Poultry Sci.* **35**, 224–225 (1956).
- Cooper, M. D., Raymond, D. A., Peterson, R. D., South, M. A. & Good, R. A. *J. Exp. Med.* **123**, 75–102 (1966).
- Owen, J. J. T., Cooper, M. D. & Raff, M. C. *Nature* **249**, 361–363 (1974).
- Köhler, G. & Milstein, C. *Nature* **256**, 495–497 (1975).
- Hozumi, N. & Tonegawa, S. *Proc. Natl Acad. Sci. USA* **73**, 3628–3632 (1976).



A miner extracts the tantalum ore coltan in the Democratic Republic of the Congo.

MINERALOGY

Painful extractions

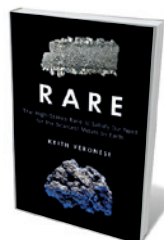
Andrew Bloodworth weighs up a study revealing the high cost of our technology-driven lust for rare metals.

Since the late 1990s, more than 5 million people have died in a civil war raging sporadically in the Democratic Republic of the Congo. This little-known, under-reported horror has been largely fuelled by conflict over minerals including coltan, an ore of tantalum — a rare metal that is a small but crucial component of almost all mobile digital devices. In 2010, a long-running dispute between China and Japan over ownership of the Senkaku Islands flared up, and China threatened to embargo shipments of rare-earth metals such as neodymium to Japan. Given China's then near-monopoly on global supply, prices of rare-earth metals spiked, and some Western governments became seriously worried about the threat to their high-tech manufacturing industries.

These events demonstrate the geopolitical consequences of our need for an ever-widening range of elements in ever larger quantities. Competition and conflict over natural resources is nothing new, but in *Rare*, Keith Veronese gives this old story a modern twist. In an engaging, eclectic but sometimes rambling journey around this complex topic,

he takes a 'cradle to grave' approach — from the stellar origins of these elements at the dawn of the Universe to their eventual fate in toxic landfills and backyard recycling sites in Africa and Asia.

Until the mid-twentieth century, only about 15 metallic elements in the periodic table had any practical use. Since then, humankind has broadened the palette of useful metals enormously, to take in 'critical', 'exotic' or 'technology' materials, including rare-earth elements, platinum-group elements, and others from antimony to zirconium. Handheld digital devices, and their almost magical appeal to billions of consumers, are driving this expansion: a smartphone can contain up to 60 elements. Other drivers include growth in



Rare: The High-Stakes Race to Satisfy Our Need for the Scarcest Metals on Earth
KEITH VERONESE
Prometheus: 2015.

renewable-energy generation (wind turbines and solar photovoltaics) and low-carbon transport (electric and hybrid cars). As *Rare* makes plain, the nuclear and defence industries also have particular requirements for exotic and scarce raw materials such as beryllium and polonium, used to initiate reactions in nuclear fuel rods.

More sinister uses of rare metals are covered in a chapter ominously entitled 'Pale Horses' (named after a 1961 Agatha Christie novel featuring thallium poisoning). The physiological effects of some metal poisons are described in gruesomely compulsive detail. Thallium sulphate, for instance — used by the infamous 'Aunt Thally', Caroline Grills, to murder four relatives in 1940s and 1950s Australia — generally triggers kidney failure. Radioactive polonium, which fatally damages the lymphatic and immune systems, was allegedly used to kill Russian dissident Alexander Litvinenko in London in 2006.

Veronese's background as a chemist comes to the fore in his fascinating description of a tortuous process — the separation and identification of many rare-earth elements by Scandinavian scientists in the eighteenth and nineteenth centuries. A peculiar consequence of this Nordic endeavour is that the small village of Ytterby in Sweden gives its name to four elements: ytterbium, yttrium, terbium and erbium. Veronese is also authoritative when describing the link between elements' fundamental chemistry and their functional use. For example, he clearly sets out the elegant subatomic relationships behind the much-sought-after magnetic properties of rare-earth metals such as neodymium. Alloyed with iron and boron, this element forms permanent magnets with very high field strengths, key to the manufacture of hard disks and wind turbines.

However, Veronese is less comfortable in unpicking the complicated mineralogy of these elements, and how they become concentrated as ore deposits. His rather jumbled explanation may leave many readers confused about the geology of rare metals. Another weakness is an incorrect use of terminology: the term 'rare-earth metals' applies not to the metals tantalum, beryllium, rhodium or niobium, but to the 15 elements from lanthanum to lutetium (atomic numbers 57 to 71), plus scandium and yttrium.

Although sporadically distributed in *Rare*, Veronese's analyses of the various factors affecting supply of these metals adds up to a good overview. He recognizes the market's role in dictating how and where they are produced, and how geopolitics can potentially interfere with supply, particularly when it is concentrated in one country (in 2012, for example, 68% of global cobalt production came from the Democratic Republic of the Congo). He describes how China's current domination of rare-earth-element production

is mainly down to its large-scale, low-cost producers undercutting smaller operations elsewhere. Perceptively, he recognizes the significant amounts of time, money, technological know-how and risk-taking required to find an economically viable geological deposit and set up a mine and extraction plant. He also notes the technical, environmental and economic limitations of obtaining the metals through recycling.

“Until the mid-twentieth century, only about 15 metallic elements in the periodic table had any practical use.”

Real and perceived threats to supply security of these metals have led to speculation and hype about resources in more exotic locations. Veronese explores new frontiers for mining, from Antarctica to the deep ocean and even the asteroid belt. This entertaining discussion is realistic about the prospects and the huge financial, technical and environmental risks associated with attempts to recover rare metals from wilder shores. However, although many of these elements are relatively rare and geologically often unevenly distributed, it is unlikely that we will run out of any of them in the near future. As Veronese points out, it is more likely that geopolitics will trigger short-term supply disruption.

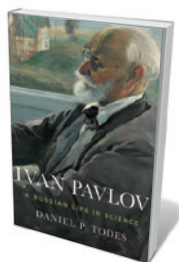


Coltan contains metals key to digital devices.

Although prone to digression and hyperbole, *Rare* makes powerful points about the consequences of conspicuous consumption. The influence of the ‘haves’ over the ‘have-nots’, the willingness of rich countries to export their environmental obligations to poor ones, and the fragility of global supply chains in a resource-constrained world are uncomfortable truths. Yet they must be told. ■

Andrew Bloodworth is science director for minerals and waste at the British Geological Survey in Nottingham, UK.
e-mail: ajbl@bgs.ac.uk

Books in brief



Ivan Pavlov: A Russian Life in Science

Daniel P. Todes OXFORD UNIVERSITY PRESS (2014)

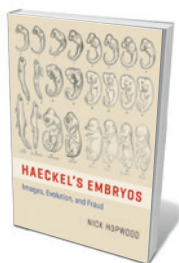
Profoundly researched, densely detailed and likely to be definitive, Daniel Todes' biography of physiologist Ivan Pavlov was 20 years in the making. The tome is a corrective, not least to the idea that Pavlov was a behaviourist: this obdurate Russian survivor of war, revolutions and Stalinization was in fact focused on “consciousness and its torments”. His immense labours in his factory-like lab produced a mixed legacy. Conditional-reflex methodology, for instance, is a staple of addiction studies, yet Pavlov's scrutiny of canine saliva for a solution to the ‘hard problem’ of neuroscience came to nothing.



Trees, Woods and Forests: A Social and Cultural History

Charles Watkins REAKTION (2015)

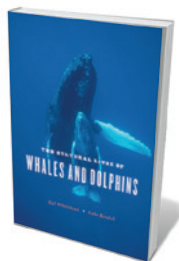
Charcoal, warships, fruit, houses, shade and sheer beauty — the manifold uses of trees have bound them inextricably to human culture. Geographer Charles Watkins' interdisciplinary exploration of that long, convoluted relationship is a fact-packed dazzler. With Watkins we walk a Neolithic ‘road’ of ash planks, delight in Pliny's description of German forests as “untouched by the ages and coeval with the world”, celebrate the rise of scientific forestry and ponder the diseases and creeping urbanization now threatening the future of these stupendous organisms. Sumptuously illustrated.



Haeckel's Embryos: Images, Evolution, and Fraud

Nick Hopwood UNIVERSITY OF CHICAGO PRESS (2015)

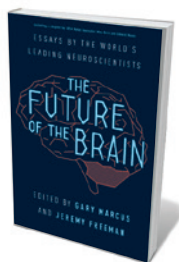
Rarely have images proved so incendiary as the embryo drawings of nineteenth-century experimental zoologist Ernst Haeckel. In this lavishly illustrated volume, Nick Hopwood traces the chequered history of the sketches, which showed similarities between embryos of higher and lower vertebrates, including humans, at particular points in their development. Haeckel intended the images as support for Charles Darwin's evolutionary theory, but under attack revealed that they were schematics. Hopwood meticulously charts how, despite the controversy, the drawings took on a life of their own.



The Cultural Lives of Whales and Dolphins

Hal Whitehead and Luke Rendell UNIVERSITY OF CHICAGO PRESS (2014)

This research round-up on cetacean culture opens with a description of one of nature's great arias: the “high sweeping squeals, low swoops, barking, and ratchets” of the humpback whale. That song, argue cetacean biologists Hal Whitehead and Luke Rendell, is the best evidence of culture in this intriguing family, because it is an indicator of social learning in action — communal singing evolves over time and changes radically over individuals' lifetimes. Fascinating findings litter this sober treatise, from sperm whales snacking off fishing longlines to the “Star Wars vocalisation” of dwarf minke.



The Future of the Brain: Essays by the World's Leading Neuroscientists

Edited by Gary Marcus and Jeremy Freeman PRINCETON UNIVERSITY PRESS (2014)

How to unify data from initiatives such as US President Barack Obama's BRAIN? In this essay compilation, editors Gary Marcus and Jeremy Freeman argue that a “confluence of new technologies” will kick-start astonishing advances in mapping, computation and simulation related to the brain. Geneticist George Church's “Rosetta Brain” sample, for instance, assembled by methods such as ‘barcoding’ cells, could prove key in brain comparison. **Barbara Kiser**



The Yellow River, seen as it passes through Haidong Prefecture in Qinghai Province, China.

RESOURCES

Silt, dams and hydraulic heroes

A study of China's Yellow River shows how water shapes the nation, finds **Philip Ball**.

China is not alone in having a water problem. But as with much else, its situation is unique. In *The Yellow River*, historian David Pietz examines why. Having done so much in modern times to expunge its past, China remains peculiarly constrained by it. Not only has every ruling power inherited the hydraulic legacy of its predecessors — for better or worse — but the idea that political legitimacy depends on “ordering the waters” is felt as keenly by modern leaders as it was by emperors from the Qin to the Qing.

Drought, flooding, unevenly distributed supplies, pollution — China suffers from them all, and then some. Floods have been recurrent and biblical, some of the worst (such as the flooding of the Yangtze and Huai in 1931) claiming more than one million lives. Droughts have produced calamitous famines, which, like that of 1876–79, could spark rumours of cannibalism.

China's water problems stem from its climate. The warm south often gets too much water; the cooler north, not enough. Both are sliced by a great river — the Yangtze to the south, the Yellow in the north — that often floods in summer. Both are fed by the glaciers of the Tibetan Plateau — the “water tower of Asia”, shrinking with global warming.

By focusing on the Yellow River and its modern management, Pietz has limited his story. But it makes sense. The Yellow has long been the most potent symbol not only of China's woes, but of its ingenuity and resilience. After epic floods and droughts in the mid-nineteenth century, it became known as “China's sorrow”; in the following century, both Nationalists and Communists cultivated a myth of nationhood in which it was the “mother of Chinese civilization”. Modern archaeology has dismantled the idea that a unified racial and cultural stock diffused from the Yellow River valley, but that story bound China's identity to its most troublesome river.

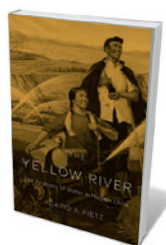
At the root of the problem is the yellow silt

that gives the river its name, which is washed in from the northern loess soil plains. The river carries more silt per litre than any other. As it settles, the riverbed rises. Containment dikes have been built up, so that much of the river is essentially an aqueduct looming over the land. Any breach unleashes a flood.

There is no easy solution. If the dikes confine the flow tightly, might its speed carry the sediment to the sea? Or is it better to let the river take its own course, confined only by levees to restrict the worst flooding? It seems impossible to find a compromise in the North China Plain, which is made from alluvial Yellow River deposits, is home to one-quarter of China's people and is the country's major source of wheat. As Pietz points out, deforestation on the loess plateau — already under way around two millennia ago — exacerbates the hazards. Attempts to tame the Yellow will always struggle unless erosion is addressed.

This book's value lies in showing that the issues are not confined to geology and engineering. Water in China is an unavoidably political issue because of cultural beliefs. Does any other nation's flood myth have a hydraulic engineer as the hero? By taming the water, Yü the Great was able to found the first dynasty, the Xia. Ever since, Pietz points out, controlling the waterways has been near synonymous with a mandate to rule.

When Sun Yat-sen founded the Republic of China in 1912, he saw that modernization depended on controlling the waters — for irrigation, flood control, transport and hydroelectric power. Turmoil, invasion and civil war hindered those plans until the



The Yellow River: The Problem of Water in Modern China

DAVID A. PIETZ
Harvard University
Press: 2015.

People's Republic was established in 1949. Mao Zedong made use of the Yü template when his propagandists insisted that only the Communist state could combine collective enterprise with technological expertise to solve the old problems. But his hydraulic-engineering programme largely failed: many big dams, such as the silt-plagued Sanmenxia on the Yellow, were poorly planned and executed, and several collapsed. As Pietz notes, some Chinese engineers admit that they are still cleaning up the mess of the Mao era.

When China's economy began its post-Mao boom, new water crises arose. Since the 1990s, damming and overuse have meant that the Yellow River has repeatedly run dry before reaching the sea, and much of the country's fresh water is too polluted for human contact.

Yet the harm to the environment caused by improper water management — from soil salinization to industrial and urban waste — shows how water remains centre stage in China's political evolution. Unwilling or unable to enforce regulations, the Chinese government has tacitly devolved some responsibility to environmental pressure groups and activists who, Pietz writes, “could be successful only by understanding the limits of their activism”, and reining in the criticism. In this way, water issues are driving a pluralism that, if not exactly democracy, is nevertheless broadening the political discourse. In one way or another, water is still shaping China. ■

Philip Ball is a writer based in London. His book *The Water Kingdom* will be out in 2016. e-mail: p.ball@btinternet.com

CORRECTION

The book review ‘Fowl Domination’ (*Nature* **515**, 490–491; 2014) wrongly called the book's author Adrian. His name is Andrew Lawler.

Carbon at the coastal interface

The extent to which coastal-ocean regions act as a sink for carbon dioxide has been enigmatic. An estimate based on more than 3 million observations suggests a smaller sink than was thought, concentrated at high latitudes.

NICOLAS GRUBER

The thin strip of coastal ocean at the interface between land and the vast areas of the open sea contains some of the most biologically productive areas of the world's oceans. These coastal regions support most of the marine resources harvested by humans¹, despite covering only a tiny fraction of the ocean's surface area. One would expect these regions to act as strong sinks for atmospheric carbon dioxide because of their high rates of photosynthetic carbon fixation. Indeed, an early estimate² from the late 1990s suggested a coastal-ocean sink strength of about 1 petagram of carbon per year (Pg C yr^{-1} ; 1 petagram is 10^{15} g) — a value that, when combined with the open-ocean sink strength³ of about $2.3 \pm 0.7 \text{ Pg C yr}^{-1}$, would make the ocean by far the largest sink for the CO_2 that humans emit every year into the atmosphere³. Reporting in *Global Biogeochemical Cycles*, Laruelle and co-authors⁴ now show that the global coastal ocean is a much smaller CO_2 sink than was thought, amounting to only about 0.2 Pg C yr^{-1} .

The authors' result continues a trend toward progressively smaller estimates, with the most recent previous publication⁵ suggesting a value between 0.2 and 0.5 Pg C yr^{-1} . But previous global estimates were based on the extrapolation of a small number of observations (often not more than a few hundred), whereas Laruelle *et al.* constrained estimates of the coastal air-sea CO_2 fluxes nearly globally using more than 3 million observations of the surface-ocean concentration of CO_2 from a recently compiled database⁶.

The authors divided the global coastal ocean into 45 segments and then determined the air-sea CO_2 flux for each segment using an approach that depended on the data density. In well-sampled regions, each segment was sub-divided into a grid, data were 'binned' into the grid's cells, and flux was computed directly on the basis of the gridded data. In areas that were sampled less well, the data were grouped according to water depth and other parameters. If the coverage was extremely limited, the researchers simply averaged the few available observations. Through this adaptive extrapolation method, they were able to constrain the air-sea CO_2 fluxes for 96% of the global coastal

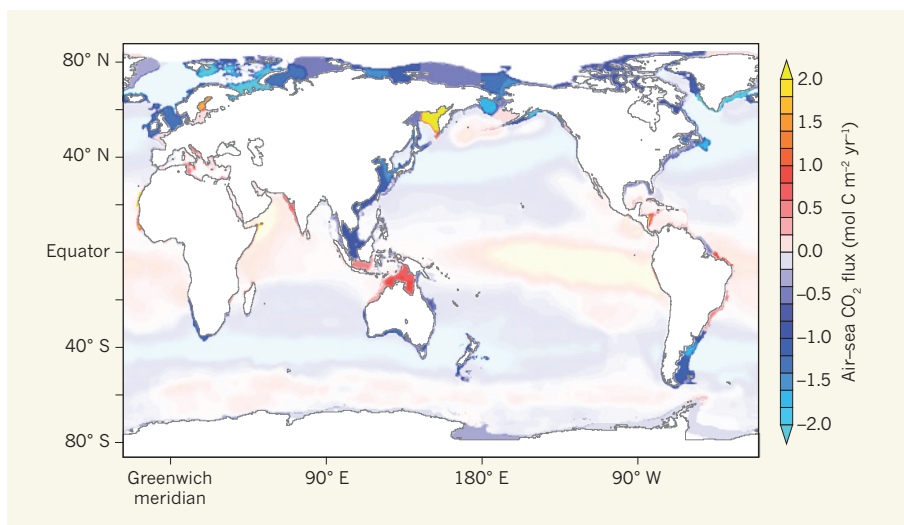


Figure 1 | Global map of air-sea CO_2 fluxes. Laruelle *et al.*⁴ report the average flux density of coastal ocean regions (strong colours) based on more than 3 million observations. Data for the open ocean¹² are shown in fainter colours, for comparison. Blue colours represent carbon sinks, other colours represent carbon sources. Units of flux are moles of carbon per square metre per year.

ocean. This great leap forward in data coverage and methodology compared with all previous studies makes this estimate by far the most solid and reliable, thus providing an important constraint for the global carbon cycle.

A striking observation arising from the coastal air-sea CO_2 fluxes is that not all regions are sinks — many are sources of CO_2 to the atmosphere (Fig. 1). The flux distribution follows largely a latitudinal trend, with coastal regions at low latitudes generally losing CO_2 to the atmosphere, whereas those at higher latitudes, and especially in the Arctic, take it up. To a zero-order approximation, this follows the distribution of CO_2 fluxes in the open ocean, where the main sources to the atmosphere are also found at low latitudes, and the largest sinks at mid- to high latitudes⁷.

At first thought, it is surprising that the distribution of CO_2 sources and sinks does not follow that of primary production — for example, there are modest differences in CO_2 flux between the coastal ocean and the adjacent open ocean, even though the magnitude of biological productivity in these two systems differs by more than fivefold. The explanation of this puzzling finding is that many processes ultimately determine whether a region is a source

or sink of atmospheric CO_2 (ref. 8). In general, the outcome depends on a complex interaction between the extent of warming or cooling (which alters the solubility of CO_2), the amount of biological activity that removes CO_2 from the system, and how much CO_2 is supplied by upwelling and mixing from deeper waters. In coastal systems, additional complexity arises from the supply of inorganic and organic carbon from rivers on the adjacent land. The general latitudinal trend observed by Laruelle and co-workers suggests that warming and cooling dominate the distribution of the air-sea CO_2 fluxes, but the many deviations from this trend remind us how crucial other processes are in modifying the coastal carbon cycle.

Despite the vast increase in the amount of data considered in the new study compared with previous reports, the spatial, and especially the temporal, coverage of the data are still quite poor. The full seasonal cycle is resolved in only a handful of regions, and spatial coverage is nearly always too coarse to properly take into account the high spatio-temporal variability that characterizes coastal systems. The uncertainty associated with the time-mean estimate provided by Laruelle *et al.* therefore remains rather large.

The data are also insufficient to assess whether there are any trends in coastal fluxes, which is a serious gap when considering that the influence of human activity on coastal systems is increasing rapidly⁹. Of particular interest is how changes in the lateral supply of carbon from land will alter the sink–source balance of coastal systems¹⁰, and how other perturbations, such as ocean acidification, deoxygenation and increasing nutrient load, will manifest themselves in the coastal carbon cycle and ultimately alter the coastal ocean's ability to take up atmospheric CO₂.

Finally, a global uptake flux of 0.2 Pg C yr⁻¹ by the coastal ocean may seem modest, but it represents an invaluable ecosystem service.

Using the current CO₂ price of about €5 (US\$6) per tonne of CO₂ in the European Union emissions-trading system¹¹, this service may be worth about €1 billion per year. ■

Nicolas Gruber is in the *Environmental Physics Group, Institute of Biogeochemistry and Pollutant Dynamics, ETH Zurich, 8092 Zurich, Switzerland.*
e-mail: nicolas.gruber@env.ethz.ch

1. Pauly, D. & Christensen, V. *Nature* **374**, 255–257 (1995).
2. Tsunogai, S., Watanabe, S. & Sato, T. *Tellus B.* **51**, 701–712 (1999).
3. Ciais, P. et al. in *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental*

- Panel on Climate Change* (Stocker, T. F. et al.) Ch. 6, 465–570 (Cambridge Univ. Press, 2013).
4. Laruelle, G. G., Lauerwald, R., Pfeil, B. & Regnier, P. *Glob. Biogeochem. Cycles* <http://dx.doi.org/10.1002/2014GB004832> (2014).
 5. Bauer, J. E. et al. *Nature* **504**, 61–70 (2013).
 6. Bakker, D. C. E. et al. *Earth Syst. Sci. Data* **6**, 69–90 (2014).
 7. Gruber, N. et al. *Glob. Biogeochem. Cycles* **23**, GB1005 (2009).
 8. Takahashi, T. et al. *Deep Sea Res. Part II Top. Stud. Oceanogr.* **56**, 554–577 (2009).
 9. Doney, S. C. *Science* **328**, 1512–1516 (2010).
 10. Regnier, P. et al. *Nature Geosci.* **6**, 597–607 (2013).
 11. Koch, N. et al. *Energy Policy* **73**, 676–685 (2014).
 12. Landschützer, P. et al. *Glob. Biogeochem. Cycles* **28**, 927–949 (2014).

This article was published online on 10 December 2014.

BIOPHYSICS

Enzymes surf the heat wave

Molecular diffusion of some enzymes is enhanced when they catalyse reactions, but the reason for this was obscure. Dissipation of heat generated by catalysis through the protein is now thought to propel the molecules. SEE LETTER P.227

A. JOSHUA WAND

Ever since 1894, when Emil Fischer proposed that enzyme substrates fit perfectly into active sites like a key in a lock¹, biochemists have been fascinated by the catalytic action of enzymes. For more than a century, the focus has been almost entirely on the details and principles that govern the interconversion of reactant and product molecules. The fate of the accompanying heat of enzymatic reactions has been largely ignored. It is from this perspective that Riedel *et al.*², in a paper on page 227 of this issue, investigate the

origin of the anomalous diffusion of several enzymes, which they ascribe to an acoustic wave generated by the heat released during catalysis.

Previous studies indicated that the hydrolysis of urea by urease³ and the conversion of hydrogen peroxide to molecular oxygen and water by catalase⁴ enhanced the molecular diffusion of these enzymes. These observations for urease were interpreted as self-phoretic effects — that is, they were thought to be generated by the release of charged products from the enzyme's surface⁵. Riedel and colleagues examined an alternative hypothesis: that dissipation of the heat of the catalysed reaction is responsible.

Using a technique known as single-molecule fluorescence correlation spectroscopy, the authors demonstrated that the diffusion of four enzymes — including urease and catalase — correlates with the rates of the reactions catalysed by those enzymes, and therefore with the heat produced. Several crucial control experiments eliminated various other potential causes, such as local solvent-heating effects or simple binding of a ligand to the enzyme. The researchers used the enzyme triose phosphate isomerase as a negative control because it has little heat of reaction, and observed no anomalous diffusion in this case.

Riedel *et al.* report that, for each enzyme, the fractional increase in diffusivity of the macromolecule is proportional to the velocity of the reaction (and so to the heat generated). There is no common proportionality, however, suggesting that the microscopic details of each protein are key. To simulate the proposed effect, the authors heated the catalytic centre of catalase using a short laser pulse and observed qualitatively the same anomalous diffusion seen during catalysed reactions. It therefore seems that the flow of heat through the protein molecule does indeed give rise to the peculiar diffusion of these enzymes.

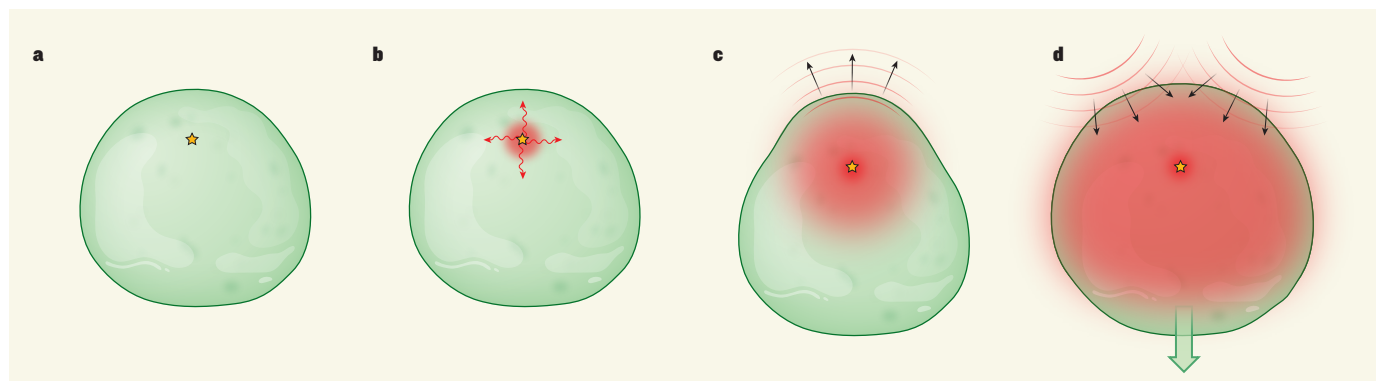


Figure 1 | The chemoacoustic model of anomalous enzyme diffusion. Riedel *et al.*² report that dissipation of heat generated during enzymatic reactions increases the diffusion of certain enzyme molecules. **a–c**, They suggest that heat released at the active site (yellow star) during a catalytic event generates a radial deformation wave that causes the enzyme to rapidly expand; orange

areas indicate passage of the wave. This causes acoustic waves (black arrows) in the surrounding solvent. **d**, If reflected back on the protein, the waves cause the molecule to move (green arrow). Because enzyme molecules also undergo rotational Brownian motion (not shown), this mechanism of locomotion will not generate overall motion in a particular direction.

But how does the heat of reaction enable the enzyme molecule to move? Here, it must be noted that proteins exist in a world in which Brownian motion is governed by viscous forces, rather than by inertia^{6,7}. Coasting is not an option — continuous force generation is required. And not all types of motion can cause translational diffusion⁶.

To explain their observations, Riedel and colleagues develop a simple model whereby the heat generated from each catalytic cycle is transmitted through the enzyme as a pressure wave (Fig. 1). In this model, the active site must be asymmetrically placed — that is, not at the enzyme's centre of mass. The pressure wave creates differential stress at the enzyme–solvent interface, which in turn propels the enzyme. The authors call this a 'chemoacoustic' effect. The model itself is barren of microscopic details and assumes that the enzyme, solvent water and their interface are individually homogeneous. This simple view nevertheless gives reasonable estimates of the force generated at the enzyme–water interface from a modest fluctuation in volume created by the thermal activation of a protein's motional modes.

Over the past few decades it has become abundantly clear that proteins (including enzymes) are dynamic entities rich in motion over a vast range of timescales. The picosecond-to-nanosecond time frame relevant to Riedel and co-workers' findings is no exception. Sophisticated molecular-dynamics simulations⁸ suggest that transmission of energy through a protein can be remarkably fast — on the order of 5 ångströms per picosecond (1 picosecond is 10^{-12} seconds) — and non-uniformly distributed. The complexity of the internal motion of protein molecules has also been exemplified using nuclear magnetic resonance spectroscopy⁹. This complexity is particularly apparent in the context of applied pressure¹⁰, which is highly relevant to the present study. But perhaps most importantly, proteins have elements (often termed 'foldons') that cooperatively fold to adopt a particular substructure and dictate many of the kinetic and thermodynamic properties of protein molecules¹¹. Given all this complexity, the precise mechanism by which pressure waves move through protein molecules remains uncertain.

What happens when the transient expansion of an enzyme arrives at the enzyme–water interface is also more complicated than the situation in Riedel and colleagues' simple model. Theory^{12,13}, simulation¹⁴ and experiment¹⁵ suggest that the interaction of a protein surface with surrounding water molecules is context-dependent and variable. Furthermore, evidence^{14,16} seems to suggest that proteins can induce long-range ordering of water, beyond the traditional 'hydration layer' of water molecules that immediately surrounds a dissolved protein molecule. Clearly there is much to do to fully understand how heat generated by

enzyme catalysis is dissipated and how this can result in a locomotive protein molecule.

Finally, the most intriguing question of all is whether the anomalous diffusion of protein molecules through the chemoacoustic effect is a product of evolution or simply an accidental unselected result of heat flow in proteins. It seems reasonable to think that enzymes might have evolved the capacity to seek 'greener pastures' of substrates. But this view is more complicated than it might at first look^{6,7}, and any selective advantage may ultimately be quite obscure. ■

A. Joshua Wand is in the Department of Biochemistry and Biophysics, University of Pennsylvania, Philadelphia 19104-6059, USA. e-mail: wand@upenn.edu

1. Fischer, E. *Ber. Dt. Chem. Ges.* **27**, 2985–2993 (1894).
2. Riedel, C. et al. *Nature* **517**, 227–230 (2015).
3. Muddana, H. S., Sengupta, S., Mallouk, T. E., Sen, A.

& Butler, P. J. *J. Am. Chem. Soc.* **132**, 2110–2111 (2010).

4. Sengupta, S. et al. *J. Am. Chem. Soc.* **135**, 1406–1414 (2013).
5. Paxton, W. F., Sundararajan, S., Mallouk, T. E. & Sen, A. *Angew. Chem. Int. Edn* **45**, 5420–5429 (2006).
6. Purcell, E. M. *Am. J. Phys.* **45**, 3–11 (1977).
7. Golestanian, R. *Phys. Rev. Lett.* **102**, 188305 (2009).
8. Sharp, K. & Skinner, J. J. *Proteins* **65**, 347–361 (2006).
9. Igumenova, T. I., Frederick, K. K. & Wand, A. J. *Chem. Rev.* **106**, 1672–1699 (2006).
10. Fu, Y. et al. *J. Am. Chem. Soc.* **134**, 8543–8550 (2012).
11. Englander, S. W. & Mayne, L. *Proc. Natl Acad. Sci. USA* **111**, 15873–15880 (2014).
12. Sharp, K. A., Nicholls, A., Fine, R. F. & Honig, B. *Science* **252**, 106–109 (1991).
13. Chandler, D. *Nature* **437**, 640–647 (2005).
14. Heyden, M. & Tobias, D. J. *Phys. Rev. Lett.* **111**, 218101 (2013).
15. Nucci, N. V., Pometun, M. S. & Wand, A. J. *Nature Struct. Mol. Biol.* **18**, 245–249 (2011).
16. Grossman, M. et al. *Nature Struct. Mol. Biol.* **18**, 1102–1108 (2011).

This article was published online on 10 December 2014.

CLIMATE SCIENCE

Unburnable fossil-fuel reserves

How much more of Earth's fossil fuels can we extract and burn in the short- to medium-term future and still avoid severe global warming? A model provides the answer, and shows where these 'unburnable' reserves are. [SEE LETTER P.187](#)

MICHAEL JAKOB & JÉRÔME HILAIRE

Cumulative carbon dioxide emissions must be less than 870 to 1,240 gigatonnes between 2011 and 2050 if we are to have a reasonable chance of limiting global warming to 2 °C above the average global temperature of pre-industrial times¹. But the carbon contained in global resources of fossil fuels is estimated² to be equivalent to about 11,000 Gt of CO₂, which means that the implementation of ambitious climate policies would lead to large proportions of reserves remaining unexploited (Fig. 1). On page 187 of this issue, McGlade and Ekins² comprehensively quantify the regional distribution of fossil-fuel reserves that should not be burned between 2010 and 2050, by modelling a broad range of scenarios based on least-cost climate policies.

Several studies have previously analysed the global long-term implications of climate-change mitigation on fossil-fuel markets^{3–5}. The novelty of the present study stems from the detailed regional representation of fossil-fuel reserves used in the authors' model, which are based on well-established data sources. In each of the 16 regions modelled, fossil fuels are divided into 21 categories that include various

types of coal, oil and gas. Each category further accounts for key characteristics, such as recoverable resources, production and trade costs, as well as natural decline rates of production (the rates of fall that would occur in the absence of any further investment).

This approach allows the authors to emphasize differences in unburnable fossil-fuel reserves. About 80%, 50% and 30% of coal, gas and oil reserves, respectively, would need to remain below Earth's surface if the world is to limit an increase in global mean temperature to 2 °C. The uneven distribution of unburnable carbon has far-reaching consequences for fossil-fuel owners.

For example, the Middle East, which holds the bulk of conventional oil reserves, would need to leave about 40% of those reserves underground. This corresponds to about 8 years of global production at current levels⁶ (87 million barrels per day). Similarly, countries with large coal endowments would face great challenges. China and India would have to discard 66% of their reserves, whereas Africa would have to leave 85% of them. In addition, the United States, Australia and countries of the former Soviet Union would need to leave more than 90% of their coal reserves underground, in stark contrast to the renaissance of coal

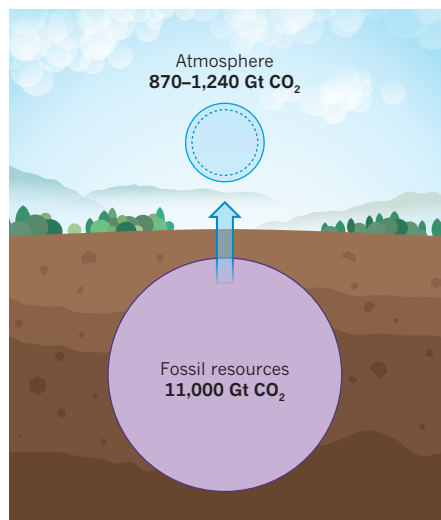


Figure 1 | Fossil-fuel resources exceed atmospheric disposal space for carbon emissions. McGlade and Ekins² report that the carbon contained in fossil-fuel reserves (equivalent to 11,000 gigatonnes of carbon dioxide) is much more than the amount that can be emitted as CO₂ to the atmosphere (870–1,240 Gt) if global warming is to be limited to 2 °C above the average global temperature of pre-industrial times. (Figure adapted from ref. 14.)

use currently under way in many places⁷.

Gas-fired power plants emit less CO₂ per unit of energy produced than coal-fired plants, and so ‘unconventional’ sources of natural gas, such as shale gas, have been touted as a bridge to the projected global transition to carbon-free, renewable energy technologies (although this bridging role has recently been challenged⁸). Encouraged by the recent shale-gas production boom in the United States, several world regions, including China, India, Africa and the Middle East, are seeking to unlock their large endowments or increase existing production. However, McGlade and Ekins’ analysis shows that Africa and the Middle East would have to leave their entire unconventional gas resources underground, and that about 10% of the combined endowment of China and India (which includes substantial amounts of coal-bed methane) could be produced.

McGlade and Ekins’ figures, computed for the period 2010–50, show that the amounts of unburnable fossil fuels are modestly sensitive to the availability of carbon capture and sequestration technology. When this technology is not available, even less coal, oil and gas can be extracted, and natural gas must be used in preference to coal because of the gas’s lower ratio of emissions to energy produced. The future use of CO₂-removal technologies might allow further extraction of all fossil fuels after 2050, but there are many uncertainties associated with predicting the availability of these young technologies.

The authors’ insights echo calls⁹ in the past

few years for society to divest itself of fossil fuels. Such calls have been made by organizations in an attempt to influence institutional investors, such as pension funds, to shift their portfolios towards clean-energy investments. These organizations also draw attention to a potential bursting of the ‘carbon bubble’ that would result from the adoption of ambitious climate policies, leading to severe devaluations of fossil-fuel reserves, which are currently worth about US\$27 trillion⁹. Fossil-fuel companies must therefore ask themselves whether they should continue to invest in exploration for, and processing of, oil, gas and coal, or risk losing billions of dollars of stranded assets. Given the political influence of the fossil-fuel industry, policy-makers must design solutions that ensure stakeholders’ acceptance.

Importantly, McGlade and Ekins’ results clearly highlight the distributional challenge of climate policy: imposing a limit on the use of fossil fuels transfers economic benefits (known as rents) from resource owners to those who obtain the right to use the remaining burnable reserves. Hence, successful climate policy will crucially hinge on the question of whether this ‘climate rent’ can be shared in an equitable way that also ensures resource owners are compensated for their losses⁴. This could be achieved by an appropriate allocation of emissions permits in an international carbon market, or by payments through the Green Climate Fund (which was set up by the United Nations to assist developing countries in adopting practices that counter climate change). Other proposals include alleviating national debt in exchange for emissions reductions¹⁰, or using some part of the climate rent to finance access to basic infrastructure services, such as water, sanitation and electricity¹¹.

But given the crucial role of energy in economic development, how can countries be convinced to forgo the use of fossil fuels if this is perceived to imperil primary policy objectives such as poverty reduction? During the US–Africa Leaders’ Summit last August, for example, Tanzania’s energy minister, Sospeter Muhongo, said¹²: “We in Africa, we should not be in the discussion of whether we should use coal or not. In my country of Tanzania, we are going to use our natural resources because we have reserves which go beyond 5 billion tons.” Only a global climate agreement that compensates losers and is perceived as equitable by all participants can impose strict limits on the use of fossil fuels in the long term. By identifying potential winners and losers of climate-change mitigation, analyses such as the one by McGlade and Ekins provide valuable support for the design of such an agreement, and inform short-term measures that can pave the way to an accord¹³. ■

Michael Jakob and Jérôme Hilaire are at the Potsdam Institute for Climate Impact



50 Years Ago

The Schizophrenia Research Fund has been established to support research into problems connected with mental illness in general and schizophrenia in particular ... Initial impetus has been given to the fund by a gift of £50,000 from the Rothschild family, and the establishment of a Schizophrenia Research Fellowship, to which Dr. D. Straughan has been appointed ... Dr. Straughan’s contract is for seven years. His work ... has been concerned with pharmacological aspects of mammalian brain physiology, and he will concentrate on the biochemical basis of schizophrenia. It is hoped that this initial effort will attract interest in, and support for, work in the immense field of research bearing on the problems of mental health.

From *Nature* 9 January 1965

100 Years Ago

There is a widespread but erroneous belief in official circles, and among wealthy philanthropists, to the effect that you can hire a scientific discoverer and then say to him, “Discover me this” or “Discover me that” (naming to him a possible and greatly desired piece of new knowledge), and that he will thereupon proceed right away to make the discovery which you want ... But valuable and important scientific discovery cannot be produced directly in response to orders given and money expended. You cannot manufacture scientific discovery like soap. The great difficulty, in the first place, is to catch that rare and evasive creature — a scientific discoverer — and when you have found him you have to humour him and let him do as he fancies. Then he will discover things, but probably not the things which either you or he wanted or expected.

From *Nature* 7 January 1915

Research, Head of Research Domain III, 14412 Potsdam, Germany. M.J. is also at the Mercator Research Institute on Global Commons and Climate Change, 10829 Berlin, Germany.

e-mails: jakob@mcc-berlin.net; hilaire@pik-potsdam.de

- Clarke, L. et al. in *Climate Change 2014: Mitigation of Climate Change. Contribution of Working Group III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change* (eds Edenhofer, O. et al.) Ch. 6 (Cambridge Univ. Press, 2014).
- McGlade, C. E. & Ekins, P. *Nature* **517**,

- 187–190 (2015).
- McCollum, D., Bauer, N., Calvin, K., Kitous, A. & Riahi, K. *Clim. Change* **123**, 413–426 (2014).
- Bauer, N. et al. *Clim. Change* <http://dx.doi.org/10.1007/s10584-013-0901-6> (2013).
- Jewell, J. et al. *Clim. Change Econ.* **04**, 1340011 (2013).
- BP Statistical Review of World Energy 2014; www.bp.com/en/global/corporate/about-bp/energy-economics/statistical-review-of-world-energy.html (2014).
- Steckel, J. C., Jakob, M., Marschinski, R. & Luderer, G. *Energy Policy* **39**, 3443–3455 (2011).
- McJeon, H. et al. *Nature* **514**, 482–485 (2014).
- Carbon Tracker & Grantham Institute *Unburnable Carbon 2013: Wasted Capital and Stranded Assets*; www.carbontracker.org/report/

- wasted-capital-and-stranded-assets (2013).
- Fenton, A., Wright, H., Afionis, S., Paavola, J. & Huq, S. *Nature Clim. Change* **4**, 650–653 (2014).
- Jakob, M. & Edenhofer, O. *Oxford Rev. Econ. Policy* (in the press).
- www.scientificamerican.com/article/africa-needs-fossil-fuels-to-end-energy-apartheid
- Jakob, M. et al. *Nature Clim. Change* **4**, 961–968 (2014).
- Edenhofer, O., Flachsland, C., Jakob, M. & Lessmann, K. *The Atmosphere as a Global Commons — Challenges for International Cooperation and Governance Discuss. Pap.* 13-58; http://belfercenter.ksg.harvard.edu/files/hpcadp58_edenhofer-flachsland-jakob-lessmann.pdf (2013).

CANCER

Resistance through repopulation

Bladder-cancer cells have been found to release prostaglandin E_2 when they are killed by chemotherapy. Paradoxically, this molecule stimulates the proliferation of surviving cancer stem cells, leading to tumour repopulation. **SEE LETTER P.209**

IAN F. TANNOCK

Initial or acquired resistance to anticancer drugs limits their effectiveness. Many papers have described mechanisms underlying such drug resistance, but most have evaluated molecular changes in single tumour cells, with the emphasis on stable genetic changes that may be induced by therapy or selected for. However, causes of drug resistance are multiple and complex, and in

this issue, Kurtova *et al.*¹ (page 209) explore an alternative mechanism — the repopulation of tumours by cancer stem cells that survive treatment. The authors show that, in certain bladder tumours, repopulation is stimulated by the hormone-like lipid molecule prostaglandin E_2 (PGE_2), which is released from tumour cells that are killed by the initial chemotherapy. Furthermore, the authors show that repopulation can be inhibited by the drug celecoxib, which inhibits PGE_2 synthesis.

As Kurtova and colleagues point out, cellular repopulation of normal tissues is crucial in wound healing and in the recovery of white blood cells of the immune system following chemotherapy. The rapidly proliferating, drug-sensitive precursor cells in bone marrow are preferentially killed by anticancer drugs, leading to a decrease in functional white cells in the blood about 10–14 days after treatment and leaving patients at risk of infection. The fall in white-blood-cell count stimulates the production of haematological growth factors, which signal slowly proliferating bone-marrow stem cells to divide, and the blood is usually repopulated with mature white cells in about three weeks, the usual interval between courses of clinical chemotherapy.

But tumour repopulation also occurs, through cancer cells that survive repeated treatments, and this can accelerate with time². It has long been recognized that repopulation occurs during courses of radiotherapy, and that the relative rates of repopulation in tumour and surrounding normal tissue can determine patient outcomes³. Kurtova and colleagues' findings add to evidence that repopulation is also a cause of treatment failure during chemotherapy. Mathematical models show² that accelerating repopulation during successive courses of treatment can lead to tumour remission then regrowth, as is observed in patients, without any change in the intrinsic sensitivity of the tumour cells to treatment drugs (Fig. 1).

Inhibiting repopulation has the potential to improve the outcome of chemotherapy, provided that the inhibitors used are specific for tumour cells and do not prevent the essential repopulation in normal tissues. Although the mechanisms leading to tumour repopulation are not known, it has been assumed that it might occur as tumours shrink and the nutrient status of surviving cells improves as other cells die². The surviving population will probably over-represent cells that were nutrient deprived and slowly proliferating, and thus resistant to drugs that target cell-cycle pathways. The surviving cells are also probably situated far from tumour blood vessels, because this location will have been subject to lower drug concentrations owing

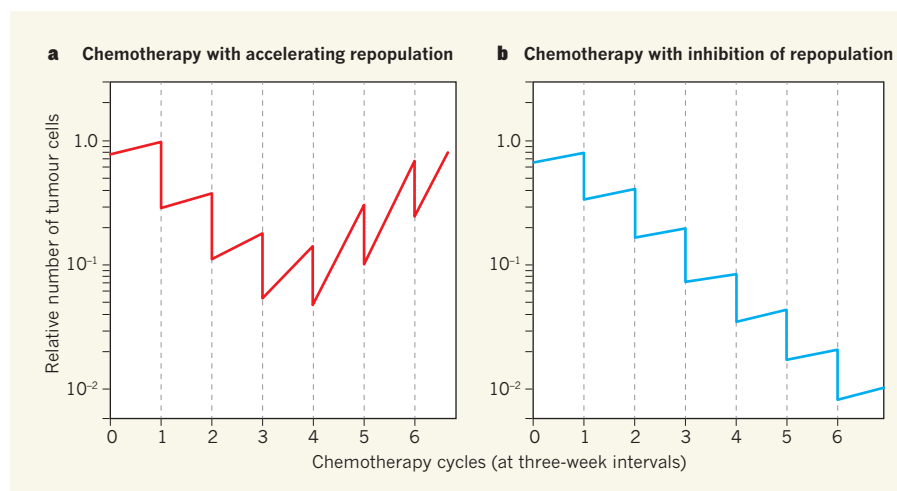


Figure 1 | Inhibiting repopulation enhances chemotherapy effects. **a**, Repopulation of a tumour with cells that survive chemotherapy can lead to patterns of tumour shrinkage and regrowth during and after each cycle of chemotherapy. This repopulation often accelerates, even if there is no change in the effectiveness of the chemotherapy in killing tumour cells during successive courses. **b**, Selective inhibition of repopulation between treatment cycles can lead to continued loss of tumour cells. Kurtova *et al.*¹ show that this approach works in mouse models of bladder cancer, by using the drug celecoxib to inhibit the production of the molecule prostaglandin E_2 , which they found stimulates repopulation from cancer stem cells in this tumour.

to the drugs' limited diffusion from those blood vessels⁴.

Kurtova and colleagues add to these previous suppositions about cancer repopulation by providing evidence that there is selective repopulation from previously slowly proliferating bladder-tumour cells that have markers suggesting that they are tumour stem cells. Furthermore, they show that the release of PGE₂ by cells undergoing programmed apoptotic death as a result of chemotherapy directly stimulates these putative tumour stem cells to proliferate. The enzyme COX2, which mediates PGE₂ production, had previously been associated with inflammation and carcinogenesis in the urinary bladder⁵. Kurtova *et al.* demonstrate that inhibiting COX2 with celecoxib inhibits tumour-cell repopulation and improves the antitumour effects of chemotherapy in mice with tumours derived from human bladder-cancer cells.

These results emphasize the importance of studying drug resistance in solid tumours. Repopulation is just one of several factors limiting the relevance of investigations that use dispersed cells in culture; others include the steep gradients of drug concentration in solid tumours, which results in inadequate drug delivery to many tumour cells, and the influence of cellular contact on the response^{4,6,7}. Kurtova and colleagues did not take these

factors into account, and, indeed, PGE₂ production is probably only one of several mechanisms by which tumour-cell repopulation is induced during chemotherapy: in different tumours, other mechanisms, both specific (for example, the activity of growth factors) and nonspecific (such as improved nutrition in the tumour microenvironment), are probably involved.

Another possible limitation of Kurtova and colleagues' study is that they treated the mice continuously with celecoxib during cycles of chemotherapy. Celecoxib inhibits the production of PGE₂, thereby removing the proliferative stimulus for tumour stem cells. But to achieve maximum benefit, cell-cycle-inhibiting drugs such as celecoxib would need to be given between cycles of chemotherapy, and to be removed before the start of the next course of (cycle-dependent) chemotherapy. Surviving tumour cells are therefore then proliferating and sensitive to the anticancer drugs.

An obvious question is whether treatment with celecoxib can also improve the outcome of chemotherapy in patients. There have been several randomized clinical trials in which celecoxib has been added to chemotherapy (although probably not in an ideal schedule) for the treatment of various cancers, although I am not aware of such trials for patients with bladder cancer. A meta-analysis of these trials shows

some increase in response rate, but no change in the more important endpoint of one-year survival⁸. Also, although inhibition of repopulation owing to PGE₂ might be expected to be specific to tumour cells, the addition of celecoxib led to increased cardiovascular toxicity and anaemia⁸. Despite these caveats, Kurtova and colleagues' findings establish the principle that treatment failure can be caused by mechanisms of drug resistance that operate only *in vivo*, but that it is possible to modify such mechanisms to improve therapeutic outcome. ■

Ian F. Tannock is in the Department of Medical Oncology, Princess Margaret Cancer Centre, Toronto, Ontario M5G 2M9, Canada. e-mail: ian.tannock@uhn.ca

1. Kurtova, A. V. *et al.* *Nature* **517**, 209–213 (2015).
2. Kim, J. J. & Tannock, I. F. *Nature Rev. Cancer* **5**, 516–525 (2005).
3. Fenwick, J. D., Pardo-Montero, J., Nahum, A. E. & Malik, Z. I. *Int. J. Radiat. Oncol. Biol. Phys.* **82**, 1021–1030 (2012).
4. Trédan, O., Galmarni, C. M., Patel, K. & Tannock, I. F. *J. Natl Cancer Inst.* **99**, 1441–1454 (2007).
5. Gakis, G. *Adv. Exp. Med. Biol.* **816**, 183–196 (2014).
6. Teicher, B. A. *et al.* *Science* **247**, 1457–1461 (1990).
7. Green, S. K., Frankel, A. & Kerbel, R. S. *Anticancer Drug Design* **14**, 153–168 (1999).
8. Chen, J. *et al.* *Clin. Ther.* **36**, 1253–1263 (2014).

This article was published online on 3 December 2014.

QUANTUM INFORMATION

Spin memories in for the long haul

Spin systems have now been found that have lifetimes of up to six hours. They could be used to build quantum-communication networks and, if optical transmission fails, could even be shipped as a 'quantum memory stick'. SEE LETTER P.177

JOHN J. L. MORTON & KLAUS MÖLNER

It is just a little past the time of year when many children write letters to Father Christmas. While being typed, these electronic letters exist in their computers' dynamic memories, before being transmitted down optical fibres to the North Pole, where they are stored on hard disks and backed up using magnetic-tape storage (or so we are reliably informed). Quantum information offers the prospect of fundamentally secure communication, but constructing practical quantum networks relies on developing a similar set of interconnected, long-lived memories storing quantum bits (qubits; the smallest logical units of quantum information). On page 177 of this issue, Zhong *et al.*¹ show how qubits encoded in the magnetic moment (spin) of

atomic nuclei in inorganic crystals can have lifetimes of up to six hours, and also offer a route to strong coupling with light. Six hours may seem short compared to storage times of conventional memories, but it should be plenty to develop quantum memories for the nodes that plug together quantum-communication networks. It is also striking that, using this technology, we could, rather straightforwardly, send a qubit from London to New York by air freight — something currently impossible using optical links.

Quantum cryptography is a way to communicate using information encoded in the states of quantum systems. It exploits the fact that measurements of quantum systems unavoidably affect their state, allowing communicating agents to validate that there is no eavesdropper on the line. Quantum states are

notoriously fragile, and hence communication of qubits typically relies on using the states of nature's fastest and least-interacting particles: photons. However, even photons are eventually scattered or absorbed, which has so far limited quantum communication based on transmission of individual photons to distances of 200 kilometres in optical fibres² and 140 km through free space³.

Extending such demonstrations to create a secure quantum network with global reach requires a quantum version of telecommunications signal repeaters to periodically boost the signal. Unfortunately, the same basic principles that make quantum cryptography resilient to eavesdropping forbid amplification of quantum signals without adding noise. Quantum repeaters therefore do not handle the actual data being transmitted — instead, their goal is to help to create high-quality quantum entanglement of two distant nodes of a network, which can be used to teleport information securely between them⁴. A quantum-repeater network consists of quantum memories, in the form of matter-based qubits, connected by short, low-loss optical channels⁵. Each quantum memory is first entangled with a light beam, and then pairs of these light beams are combined to entangle the memories, as was demonstrated in 2013 using 2 qubits encoded in the electron spin of nitrogen-based impurities in diamond crystals separated by 3 metres⁶. Pairs of entangled quantum memories can be

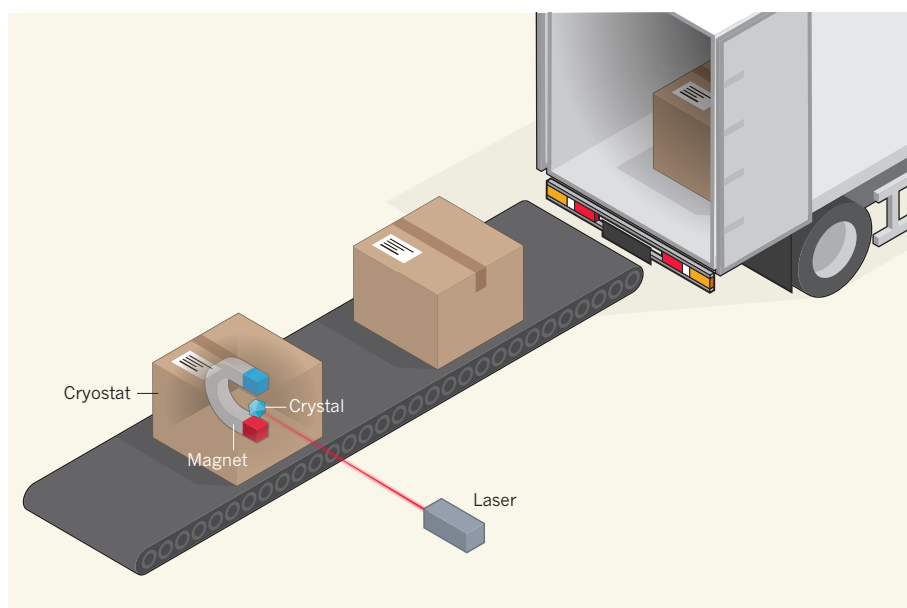


Figure 1 | Quantum postal service. Quantum bits (qubits) of information can be optically written, using laser light, into collective excitations of the nuclear magnetic moment (spin) of atomic ensembles in crystals such as yttrium orthosilicate (YSO). Zhong *et al.*¹ have shown that europium atoms in YSO can retain their nuclear spin states for as long as 6 hours at low temperatures (2 kelvin) and at a judiciously chosen magnetic field (around 1 tesla). Cryostats to keep systems at such low temperature, and magnets to produce such magnetic fields, are routinely shipped while operational, often by air, and are compact enough to fit in a modest van. On the basis of today's technology, the best way to send quantum information over long distances without loss could therefore be by post.

'daisy-chained' to form an extended quantum-repeater network, but practical quantum repeaters require the matter-based qubits to have stronger coupling to light and longer quantum lifetimes than have so far been achieved using diamond⁶.

Rare-earth elements (those with atomic numbers 57 to 71) underpin today's optical communication technologies through devices such as optical-fibre amplifiers based on erbium impurities. They may well be just as crucial in quantum networks, as demonstrated by the entanglement between light and neodymium atoms in an yttrium orthosilicate (YSO) crystal⁷. Praseodymium nuclear spins in YSO have been placed in certain 'sweet spots' of an applied magnetic field, where the spins become insensitive, to first order, to fluctuations in magnetic field that would otherwise corrupt their state. When combined with dynamical decoupling — regularly flipping the spins to cancel out low-frequency noise — such nuclear spins showed quantum lifetimes of up to half a minute at cryogenic temperatures⁸.

To go further, Zhong and colleagues turned to another rare-earth atom, europium, 'doped' in YSO, and identified conditions in which even the second-order sensitivity of the spins to magnetic-field fluctuations is suppressed, yielding predictions of quantum lifetimes of up to several minutes. The authors were in for a surprise, however, because, rather than minutes, they observed lifetimes extending up to 6 hours at temperatures of about

2 kelvin. This is because, in addition to securing the sweet-spot resilience to magnetic-field noise, the large applied magnetic field (1 tesla) also suppressed environmental fluctuations in the immediate vicinity of the europium atom, offering a double whammy of protection.

Although nuclear spins in YSO have extraordinary potential as quantum memories, more work is needed to 'write' qubits into them using optical means. There are many methods for this. In one scheme, an optical transition between two energy states in an ensemble of rare-earth atoms is driven with a laser beam, while a detector awaits the arrival of a single photon emitted by the ensemble at a wavelength or direction differing from that of the laser beam. When this happens, a particular pattern of quantum excitation becomes imprinted in the ensemble of nuclear spins (for example, imagine that the spins in the crystal adopt a striped pattern). Now, if we instead look for a photon in the interference pattern of light emitted from two such ensembles, it is possible to imprint entangled states across the two crystals (now the spins are striped in one crystal and plain in the other, and vice versa, like two pairs of Christmas socks that got mixed up before being wrapped).

Another challenge for quantum repeaters is dealing with errors, such as photon loss, which can be mitigated using a register of several coupled qubits located at each matter-based node. These allow multiple poor-quality, entangled

states across distant nodes to be transformed into one high-quality, entangled state through a process known as entanglement distillation⁹. The nuclear spin 5/2 of the europium atom has 6 energy levels and so provides complexity beyond the two states of a simple qubit that could be used for such distillation.

In addition to laying the foundation for optical quantum repeaters, these long lifetimes raise the possibility of an intriguing alternative to qubit transmission: physically transporting the qubit-containing crystal instead of using optical fibres to send the qubits. Certainly, in 6 hours it is possible to travel much further than the 200 km that currently limits optical quantum communication — and sending such crystals by post would easily be the highest-fidelity qubit-transmission technology available today (Fig. 1). However, this is unlikely to become a widespread method for quantum communication once quantum repeaters become fully realized, for the same reasons of bandwidth and speed that underlie the popularity of e-mail over a posted letter. A more likely application of portable, spin-based quantum memories would be to fit them to vehicles and use them to hold on to quantum information for limited periods when an optical quantum-communication link becomes broken.

What about other memory applications, such as quantum money or quantum passports, whose serial numbers would exploit the 'unclonability' of quantum states to prevent forgeries? These uses typically require room-temperature operation and quantum lifetimes measured in months and years — but the longest room-temperature lifetimes observed so far have been around 40 minutes, for nuclear spins in silicon¹⁰. Border agencies may start worrying now about the international regulations (and taxation) for transporting quantum information between countries, but they are unlikely to need to search your hand luggage for qubits just yet. ■

John J. L. Morton is in the London Centre for Nanotechnology and Department of Electronic and Electrical Engineering, University College London, London WC1H 0AH, UK.

Klaus Mølmer is in the Department of Physics and Astronomy, Aarhus University, 8000 Aarhus C, Denmark.

e-mail: jjl.morton@ucl.ac.uk

1. Zhong, M. *et al.* *Nature* **517**, 177–180 (2015).
2. Tang, Y.-L. *et al.* *Phys. Rev. Lett.* **113**, 190501 (2013).
3. Ma, X.-S. *et al.* *Nature* **489**, 269–273 (2012).
4. Simon, C. *et al.* *Eur. Phys. J. D* **58**, 1–22 (2010).
5. Atatüre, M. & Morton, J. J. L. *Science* **345**, 510–511 (2014).
6. Bernein, H. *et al.* *Nature* **497**, 86–90 (2013).
7. Clausen, C. *et al.* *Nature* **469**, 508–511 (2011).
8. Fraval, E., Sellars, M. J. & Longdell, J. J. *Phys. Rev. Lett.* **95**, 030506 (2005).
9. Sangouard, N. *et al.* *Phys. Rev. A* **77**, 062301 (2008).
10. Saeedi, K. *et al.* *Science* **342**, 830–833 (2013).

REGENERATIVE BIOLOGY

The versatile and plastic liver

There is conflicting evidence about which cell type is responsible for liver regeneration following damage. It emerges that duct-like progenitor cells arise from hepatocytes after liver damage, a finding that reconciles previous data.

MERITXELL HUCH

Prometheus, a Titan in Greek mythology, is said to have been condemned by Zeus to have his liver plucked out by an eagle every day for eternity, only to have the organ grow back each night. In real life, as in the myth, the liver has remarkable regenerative potential, but the precise mechanism by which liver cells repopulate the tissue following damage remains unknown. Writing in *Cell Stem Cell*, Tarlow *et al.*¹ report that, following liver damage, transplanted liver cells called hepatocytes are converted into another liver cell type, the ductal progenitor cell, which then proliferates and differentiates back into a functional hepatocyte to replenish the tissue.

The liver is the largest internal organ in the human body, and is responsible for many metabolic and detoxification activities. It is composed mainly of bile-duct cells and hepatocytes, which work in conjunction with other, less-populous cell types. But the cell type responsible for repopulation of the tissue after damage has been an area of active debate.

Some studies^{2,3} in mice have indicated that hepatocytes are the major drivers of liver regeneration. However, other studies in mice⁴, rats⁵ and zebrafish⁶ have found

that, following blockade of hepatocyte regeneration, damage activates a response that involves progenitor cells⁷. These liver progenitor cells have been shown to derive from ductal cells⁷ or to arise from an as-yet-unidentified cell^{8,9}. To add complexity to this already conflicting set of reports, a stream of bile-duct cells with progenitor features is seen in almost all cases of human liver disease that involve cell loss¹⁰. Tarlow and colleagues sought to shed light on the issue by combining two gold-standard methods to assess stem-cell potential: cell transplantation and lineage labelling. In the latter technique, cell types are indelibly labelled and their descendants traced, even if they give rise to different cell types.

The authors first addressed the question of whether hepatocytes are the source of the progenitor cells that replenish damaged livers by using mice lacking the *Fah* gene, which encodes fumerylacetoacetate hydrolase. Hepatocytes lacking this enzyme accumulate intermediate metabolites and undergo programmed cell death. Because of this, transplanted wild-type hepatocytes have a selective advantage over host hepatocytes, and so can engraft — that is, successfully integrate into the host tissue — and repopulate the liver.

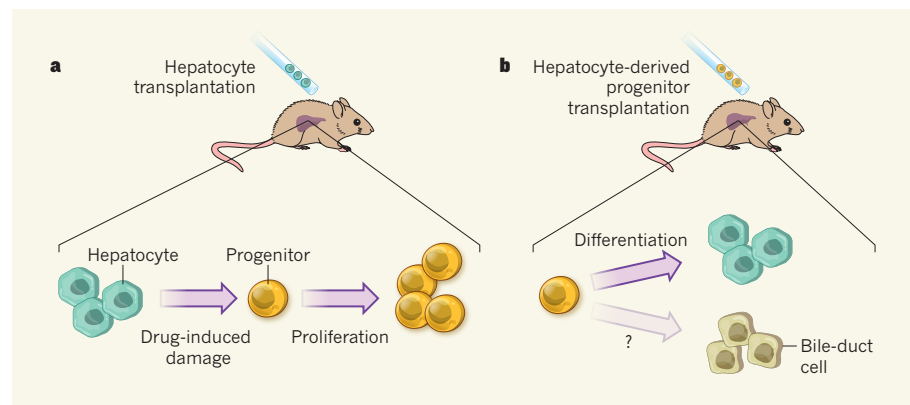


Figure 1 | Regeneration of the liver. The cell type that promotes liver regeneration following damage is a subject of much debate. **a**, Tarlow *et al.*¹ transplanted liver cells called hepatocytes into mice that were deficient in the enzyme fumerylacetoacetate hydrolase and so model liver disease. The authors then damaged the host liver and observed that the transplanted donor hepatocytes became progenitor cells with the characteristics of bile-duct-cell progenitors, which proliferated and replenished the liver. **b**, If these hepatocyte-derived, duct-like progenitor cells are taken from the host liver and engrafted into a second mouse, they can differentiate back into functionally mature hepatocytes and populate the liver. Whether they can also produce mature bile-duct cells is unknown.

Tarlow and colleagues transplanted hepatocytes labelled with a red fluorescent protein into *Fah*-deficient mice. Once the host liver had been repopulated with the red-labelled donor cells, the authors used toxic drugs to induce a type of liver injury that prevents hepatocyte proliferation and leads to the production of duct-like progenitor cells that repopulate the damaged area. They found that many of the duct-like progenitors were labelled red, indicating that they had arisen from the transplanted hepatocytes, in agreement with previous results in rats¹¹. Then the authors isolated these duct-like progenitors from the damaged livers and showed that the cells could engraft and repopulate a liver in a second animal, giving rise to fully mature hepatocytes (Fig. 1). These results indicate that hepatocyte-derived, duct-like progenitors are at least partially responsible for repopulating the liver following damage.

The experiments raise questions about whether conversion between different cell states happens in undamaged livers, and whether transformation from hepatocytes to duct-like progenitors and back again is a common phenomenon in human liver disease. Tarlow *et al.* observed a similar conversion of hepatocytes into duct-like progenitor cells when they transplanted human hepatocytes into *Fah*-deficient mice, confirming observations from tissues derived from people with acute liver failure¹².

It is not known whether duct-like progenitors from humans or mice can convert into mature bile-duct cells because of a lack of markers to distinguish donor duct cells from those derived from the mouse host. One possible way around the problem would be to grow progenitor duct cells in culture, then fluorescently label and transplant the cells. This strategy has been successfully used in mice in a study in which duct-like progenitor cells that had been grown *in vitro* engrafted into the host mouse liver after transplantation, albeit at low efficiency⁹.

Not all of the duct-like progenitor cells that arose following damage were labelled red, indicating that some were not derived from hepatocytes. Tarlow and co-workers did not assess the fate of this cell population. Do these cells make fully mature ducts after their engraftment? In humans, primary biliary cirrhosis and biliary atresia, which affect the biliary-duct system, are among the most common reasons both for liver transplants and for transplant rejection¹³. Therefore, identifying cells that can repopulate the bile-duct compartment is a priority. Could duct-like progenitors — those either derived from or independent of hepatocytes — be engrafted into mouse models of biliary diseases? Because *Fah*-deficient mice have a hepatocyte defect that does not allow duct engraftment, animal models of biliary disease will be needed to answer this question.

Part of the controversy over liver

regeneration has now been reconciled. Tarlow and colleagues' finding that drug-induced damage causes hepatocytes to give rise to duct-like progenitors that then differentiate back into hepatocytes explains much of the conflicting evidence reported. Maybe Proteus, the Greek sea god who could mutate into different forms, is a more appropriate mythological reference for the versatile liver than Prometheus. ■

Meritxell Huch is at the Wellcome Trust/MRC Stem Cell Institute, in the Department of Physiology, Development and Neuroscience, and at the Wellcome Trust/Cancer Research UK Gurdon Institute, University of Cambridge, Cambridge CB2 1QN, UK.
e-mail: m.huch@gurdon.cam.ac.uk

NEUROSCIENCE

A three-dimensional neural compass

The discovery that the neural navigation system of the mammalian brain acts in three dimensions sheds light on how mammals orient themselves in complex environments. [SEE ARTICLE P159](#)

DAVID C. ROWLAND & MAY-BRITT MOSER

All mammals face the challenge of navigating in complex, three-dimensional (3D) environments, whether they are swinging from branch-to-branch in forests or burrowing underground tunnels. How does the brain maintain a sense of place and direction in 3D? In a beautiful study on page 159 of this issue, Finkelstein *et al.*¹ report that bats have an internal neural compass that tracks direction in 3D during both surface locomotion and flight.

The hippocampal-parahippocampal region of the brain contains four neural cell types that together make up the core of the mammalian navigational system: place, grid, border and head-direction cells. Place, grid and border cells provide information about position, distance and the geometry of the surrounding environment, respectively. Head-direction cells provide information about bearing in the environment — they are considered to be the brain's compass. The classical head-direction cell, discovered in 1984 in the dorsal presubiculum of the rat parahippocampal region^{2,3}, responds whenever the animal faces a particular direction in the horizontal plane (an azimuth).

Finkelstein *et al.* first recorded neural activity in the dorsal presubiculum of bats as the

1. Tarlow, B. D. *et al.* *Cell Stem Cell* **15**, 605–618 (2014).
2. Schaub, J. R., Malato, Y., Gormond, C. & Willenbring, H. *Cell Rep.* **8**, 933–939 (2014).
3. Yanger, K. *et al.* *Cell Stem Cell* **15**, 340–349 (2014).
4. Fausto, N., Campbell, J. S. & Riehle, K. J. *Hepatology* **43**, S45–S53 (2006).
5. Everts, R. P., Nagy, P., Nakatsukasa, H., Marsden, E. & Thorgeirsson, S. S. *Cancer Res.* **49**, 1541–1547 (1989).
6. Choi, T.-Y., Ninov, N., Stainier, D. Y. R. & Shin, D. *Gastroenterology* **146**, 776–788 (2014).
7. Boulter, L., Lu, W.-Y. & Forbes, S. J. *J. Clin. Invest.* **123**, 1867–1873 (2013).
8. Shin, S. *et al.* *Genes Dev.* **25**, 1185–1192 (2011).
9. Huch, M. *et al.* *Nature* **494**, 247–250 (2013).
10. Gouw, A. S. H., Clouston, A. D. & Theise, N. D. *Hepatology* **54**, 1853–1863 (2011).
11. Michalopoulos, G. K., Barua, L. & Bowen, W. C. *Hepatology* **41**, 535–544 (2005).
12. Hattoum, A., Rubin, E., Orr, A. & Michalopoulos, G. K. *Hum. Pathol.* **44**, 743–749 (2013).
13. Vilarinho, S. & Lifton, R. P. *Cell* **150**, 1096–1099 (2012).

animals crawled upright on flat surfaces. The authors used a standard recording protocol for identifying head-direction cells in rats, but with one major technical advance — a tracking device that allowed them to resolve the orientation of the head in 3D. As with rats^{2,3}, they found many cells that responded to the animal's orientation in azimuth, but they also found cell types that responded to pitch (vertical orientation), roll and combinations of two or three axes. Relatively few cells responded to roll, and, perhaps not coincidentally, bats rarely rotate in the roll axis during flight. Therefore, the researchers focused primarily on azimuth and pitch.

Next, the authors inverted the bats. To understand why this is important, consider a typical landing manoeuvre in which the bat begins flying towards the eastern side of a cave, inverts itself and lands facing west (Fig. 1a). If head-direction cells merely respond to the orientation of the animal with respect to external landmarks, then when the bat passes through the point at which its direction in the horizontal plane changes from east to west, the population of active cells would instantly switch from those representing east to those representing west, resulting in an unstable cellular network. Surprisingly, Finkelstein and co-workers discovered that a cell that was tuned to the east when the animal was upright

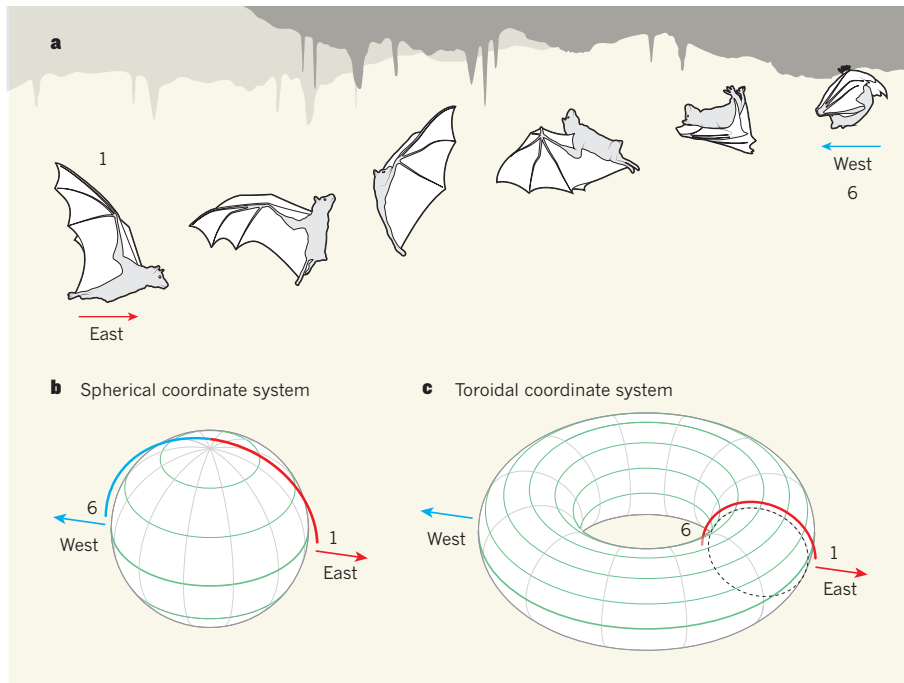


Figure 1 | Orientation during acrobatic flight manoeuvres. **a**, A typical bat landing manoeuvre, split into six stages. The bat begins facing east, inverts itself and lands facing west. **b**, Head-direction cells are the bat's internal compass. If these cells act as a spherical coordinate system, a subset of cells will respond to the direction in which the bat faces in the horizontal plane, regardless of its pitch. In this example, the bat's head changes direction in the horizontal plane (green rings) as it passes from stage 3 to stage 4, and so the population of active horizontally tuned cells would instantly switch from east-preferring (red) cells to west-preferring (blue) cells. This would make the network of cells unstable. **c**, Finkelstein *et al.*¹ propose that the bat's brain uses a toroidal coordinate system, which provides a stable representation of direction. In this system, the cells responding to eastward movement continue to respond during the entire manoeuvre. Additional cells tuned to pitch can signal whether the animal is upright or inverted. This combination gives a precise orientation in three dimensions.

was west-tuned when the animal was inverted. Thus, azimuth-coding cells tuned to one direction would remain active throughout the previously mentioned landing manoeuvre, maintaining a stable representation of direction.

However, in avoiding one problem, another arises. How can an animal gauge its orientation if azimuth-coding cells respond to one direction when it is upright and to the opposite direction when it is inverted? Finkelstein and colleagues found that the population of pitch-tuned cells cover the entire 360° range of pitch. Therefore, the intersection of active pitch-tuned and azimuth-tuned cells gives a precise orientation in 3D, and also tells the bat whether it is upright or inverted.

Taken together, the authors' data suggest that the bat's brain uses a toroidal (doughnut-shaped) coordinate system to orient itself. This system has a major advantage over a spherical coordinate system — which is perhaps more intuitive — because it is continuous over the full 360° × 360° range of pitch and azimuth. By contrast, in a spherical coordinate system, in which azimuth-tuned cells fire in a particular direction regardless of pitch, there are abrupt, destabilizing transitions as the animal passes through each pole (Fig. 1b, c).

Representation of azimuth in head-direction

cells is often modelled as a continuous network, in which connections are strongest between cells activated in similar orientations. This can be imagined as a ring of connected cells that together represent the full 360° of azimuth. Thus, as the animal rotates in the horizontal plane, neural activity moves gradually around the ring, maintaining a stable representation of direction. Extending this model to a 3D toroidal coordinate system requires only the addition of a second ring representing 360° of pitch. This probably involves connections between cells that follow a toroidal arrangement. Curiously, the same type of connectivity might underlie the firing patterns of grid cells⁴, which coexist with classical head-direction cells in the presubiculum and its two neighbouring structures, the parasubiculum and the medial entorhinal cortex⁵.

How do these results apply to other species? The acrobatic manoeuvres of bats pose challenges that do not exist for most mammals. Thus, each animal might have a neural compass that suits its own behavioural repertoire. Indeed, head-direction cells in the rat anterior dorsal thalamus (a structure that projects to the presubiculum) seem to lose their tuning when the animals are inverted⁶. Alternatively, because many mammals — including some rodents and primates — spend time hanging

upside down, many or all mammals might have the same neural compass as bats. To test this, it will be necessary to apply Finkelstein and co-workers' template to other species. For instance, the authors found 3D head-direction cells primarily in one subsection of the pre-subiculum. It will be necessary to record from this subsection in other species, and to test how these cells respond to manipulations, including inversions.

This work also has implications for other cell types in the hippocampal–parahippocampal region. Previously, the same group described 3D place cells in the bat hippocampus⁷, but the source of 3D spatial input to the hippocampus remains unclear. One plausible scenario is that this spatial information comes from the summation of grid cells encoding distances of different scales⁸, but whether grid cells process information in 3D is unknown. Because the firing patterns of grid cells probably depend on input from head-direction cells⁹, the current work strongly suggests that at least a subpopulation of grid cells will also act in 3D, making these cells attractive candidates for providing 3D spatial input to the hippocampus.

Finkelstein and colleagues' study demonstrates the immeasurable value of taking an ethological approach to neuroscience — by studying an animal that behaves in 3D, they have discovered one way in which the mammalian brain orients in 3D. This finding will undoubtedly prompt a careful re-examination of head-direction cells in rodents, and fresh investigations in other species. Other mammals may use different compasses¹⁰, and if they do, understanding what drives the differences between species (whether they are hard-wired at birth or develop with experience) will lead to insight into the basic principles of the system. ■

David C. Rowland and May-Britt Moser are at the Kavli Institute for Systems Neuroscience and at the Centre for Neural Computation, 7491 Trondheim, Norway.
e-mails: may-britt.moser@ntnu.no;
daverowland22@gmail.com

1. Finkelstein, A. *et al.* *Nature* **517**, 159–164 (2015).
2. Taube, J. S., Muller, R. U. & Ranck, J. B. Jr *J. Neurosci.* **10**, 420–435 (1990).
3. Ranck, J. in *Electrical Activity of the Archicortex* (eds Vanderwolf, G. & Buzsáki, C. H.) 217–220 (Akadémiai Kiadó, 1985).
4. McNaughton, B. L., Battaglia, F. P., Jensen, O., Moser, E. I. & Moser, M.-B. *Nature Rev. Neurosci.* **7**, 663–678 (2006).
5. Boccara, C. N. *et al.* *Nature Neurosci.* **13**, 987–994 (2010).
6. Calton, J. L. & Taube, J. S. *J. Neurosci.* **25**, 2420–2428 (2005).
7. Yartsev, M. M. & Ulanovsky, N. *Science* **340**, 367–372 (2013).
8. Solstad, T., Moser, E. I. & Einevoll, G. T. *Hippocampus* **16**, 1026–1031 (2006).
9. Giocomo, L. M., Moser, M.-B. & Moser, E. I. *Neuron* **71**, 589–603 (2011).
10. Jeffery, K. J., Jovalekic, A., Verriotti, M. & Hayman, R. *Behav. Brain Sci.* **36**, 523–543 (2013).

This article was published online on 3 December 2014.

Three-dimensional head-direction coding in the bat brain

Arseny Finkelstein^{1*}, Dori Derdikman^{1,2*}, Alon Rubin¹, Jakob N. Foerster^{1†}, Liora Las¹ & Nachum Ulanovsky¹

Navigation requires a sense of direction ('compass'), which in mammals is thought to be provided by head-direction cells, neurons that discharge when the animal's head points to a specific azimuth. However, it remains unclear whether a three-dimensional (3D) compass exists in the brain. Here we conducted neural recordings in bats, mammals well-adapted to 3D spatial behaviours, and found head-direction cells tuned to azimuth, pitch or roll, or to conjunctive combinations of 3D angles, in both crawling and flying bats. Head-direction cells were organized along a functional-anatomical gradient in the presubiculum, transitioning from 2D to 3D representations. In inverted bats, the azimuth-tuning of neurons shifted by 180°, suggesting that 3D head direction is represented in azimuth × pitch toroidal coordinates. Consistent with our toroidal model, pitch-cell tuning was unimodal, circular, and continuous within the available 360° of pitch. Taken together, these results demonstrate a 3D head-direction mechanism in mammals, which could support navigation in 3D space.

Navigation requires the knowledge of one's location ('map')^{1–3} and direction in space ('compass')^{4,5}. The neural representation of spatial location in mammals includes place-cells which compute position^{1,6–12}, together with grid cells that encode distances^{13–17}, and border/boundary cells which indicate the edges of the environment^{18–20}. The compass information is provided by head-direction cells, originally described in the rat dorsal presubiculum^{21,22}, which are neurons that become active whenever the animal's head points to a specific direction (azimuth) in the horizontal plane^{4,15,21–28}.

Orientation in 3D space is crucial for the survival of rats²⁹ and many other mammals that move in 3D space, including monkeys that jump across tree-tops and humans navigating through multi-storey buildings, as well as squirrels, cats, dolphins and bats³⁰. Yet, it remains unknown how information about the direction of an animal in 3D is represented in the brain. Previous studies investigated the activity of place-, grid- and head-direction cells in 3D space^{31–36}. However, head-direction tuning was studied primarily in planar horizontal angles (azimuth)^{15,16,21–28,37} or during locomotion on planar vertical walls^{32,34,38}, whereas the neural representation of all 3 Euler angles of head direction has not been studied, in any species. Several pitch-sensitive cells were reported in the lateral mammillary nuclei³⁹ and dorsal tegmental nucleus⁴⁰ of the rat, but they responded only to upward pitch (towards the zenith), and there seemed to be no neurons encoding directions 'straight' or 'downwards'^{39,40}. Moreover, neurons tuned to head roll were never found.

Finally, head-direction cells were suggested to provide the compass information that is thought to be essential for map representation in the entorhinal-hippocampal system^{4,41–44}. Therefore, to elucidate the neural basis of 3D navigation and in particular, how the vertical dimension is integrated into the map representation⁴⁵, it is essential to understand the coding of 3D head direction in the brain.

Neural tuning to head azimuth, pitch, and roll

Here, we developed a custom tracking device consisting of a 3D arrangement of light-emitting diodes (LEDs), which allowed monitoring all three Euler angles of head direction in freely moving animals (Fig. 1a, b

and Extended Data Fig. 1a–c; Methods). We used this device to record neural activity from the dorsal presubiculum of Egyptian fruit bats that crawled in search of food (Fig. 1b–d and Extended Data Fig. 1d–j). We conducted two behavioural sessions separated by 1 h, and analysed only neurons that showed stable directional tuning between sessions 1 and 2 ($n = 122$ directionally stable cells out of 266 total cells recorded from 4 bats; Methods). Of these 122 neurons, 78 cells (or 64%) were statistically significant head-direction cells, tuned to at least one of the three Euler angles (Methods).

We first asked whether 'classical' head-direction cells that encode azimuth are found in bat presubiculum. A large fraction of presubicular cells discharged selectively whenever the animal's head was pointing at a certain azimuth (Fig. 1e, top row). A total of $n = 63$ of the 122 cells (52%) were significantly tuned to azimuth (Methods). These azimuth-coding cells exhibited stable tuning across sessions (mean $r = 0.62$), despite their relatively low firing rates (mean peak firing rate = 1.7 Hz). The average full-tuning-width of these cells was $156^\circ \pm 6^\circ$, which was wider than that of head-direction cells in rat dorsal presubiculum^{15,21} but similar to tuning-widths reported in rat lateral mammillary nuclei³⁹. Despite these differences, other properties of bat azimuth-tuned cells were shared with rat head-direction cells: (i) the preferred head directions spanned all possible azimuth values across the neuronal population (Fig. 1f, top); (ii) preferred directions were maintained when the animals were moved passively by the experimenter (Extended Data Fig. 2a, b, f); (iii) head-direction cells exhibited very little place-tuning (Extended Data Fig. 3a–e); (iv) some head-direction cells were modulated by angular velocity (Extended Data Fig. 3f–h). Overall, despite the differences in firing rates and tuning widths, most properties of azimuthal head-direction cells in bats were very similar to classical head-direction cells in rats^{4,21,22}.

We next asked whether bat presubiculum also contained neurons tuned to pitch or roll. We found, indeed, 'pitch cells' tuned to either positive, zero or negative pitch of the head (Fig. 1e, middle), as well as 'roll cells' tuned to either right, zero or left roll of the head (Fig. 1e, bottom). Overall, 26 of the 122 cells (21%) were significantly tuned to

¹Department of Neurobiology, Weizmann Institute of Science, Rehovot 76100, Israel. ²Rappaport Faculty of Medicine and Research Institute, Technion – Israel Institute of Technology, Haifa 31096, Israel.

[†]Present address: Department of Engineering, University of Cambridge, Cambridge CB2 1PZ, UK.

*These authors contributed equally to this work.

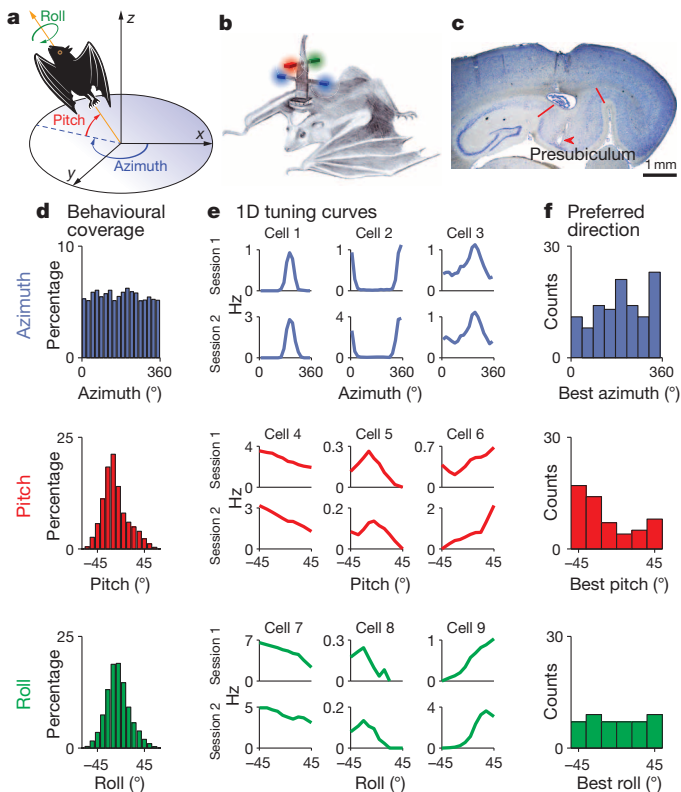


Figure 1 | Representation of head direction in the bat presubiculum: one-dimensional tuning to azimuth, pitch, and roll. **a**, Definitions of Euler angles (azimuth, pitch, roll) corresponding to the 3D rotation angles of the head. **b**, Schematic drawing of the tracking head-stage that allowed precise measurements of all 3 Euler angles of the head (see Extended Data Figs 11 and 12 and Methods for details). Illustration by S. Kaufman. **c**, Sagittal section showing tetrode track in dorsal presubiculum; arrowhead, electrolytic lesion at track end. Red lines, presubiculum borders. **d**, Behavioural distribution of head-direction angles across all upright behavioural sessions in all bats. **e**, Neural tuning curves (firing rate versus angle). Individual neurons encoded azimuth, pitch, or roll, exhibiting stable directional selectivity across behavioural sessions (rows). **f**, Preferred firing directions across the cell population span the behavioural range. Counts denote significant sessions.

pitch, and a smaller fraction, 15 cells (12%), were tuned to roll (Methods). Taken together, we identified here for the first time, to our knowledge, mammalian neurons encoding the three Euler rotation-angles of the head.

Conjunctive tuning to 3D head direction

To test whether head-direction cells encode conjunctive combinations of two Euler angles, we constructed 2D and 3D firing-rate maps in head-direction space, showing the activity of individual neurons as a function of combinations of two or three Euler angles (Fig. 2a–f). We found a diversity of directional properties in bat presubiculum neurons, including ‘pure’ cells tuned to only one Euler angle (Fig. 2b, c), but also many conjunctive cells that were significantly tuned to combinations of Euler angles (Fig. 2d, e; Methods). Double-conjunctive cells were quite common: of the ‘classical’ head-direction cells, tuned to azimuth, 27% were conjunctively tuned to pitch and 11% conjunctively tuned to roll (Fig. 2g). Triple-conjunctive neurons were rare ($n = 4$; for example, the cell in Fig. 2f, right). Thus, head direction in bat presubiculum was encoded both by neurons tuned to single Euler angles, and by multi-dimensional conjunctive neurons.

Anatomical gradient of head-direction cells

Next, we asked what is the functional–anatomical organization of head-direction cells in the dorsal presubiculum (Fig. 2h). Pure azimuth cells

were very abundant in the anterior part of presubiculum, close to the subiculum, and their prevalence decreased with distance from the subiculum border, when measured along the anterior–posterior axis (Fig. 2h and Extended Data Fig. 4b; $r = -0.97$). Opposite gradients existed for pure pitch, pure roll and conjunctive cells (Extended Data Fig. 4c–e; $r > 0.7$ in all cases). Further, for the two most common cell types, pure azimuth and conjunctive cells, we found that their abundances followed in fact a diagonal gradient along the transverse axis of dorsal presubiculum, with pure azimuth cells being most common in the antero-lateral (proximal) part of presubiculum and conjunctive cells in the postero-medial (distal) part (Fig. 2i and Extended Data Fig. 4g–o). This striking functional gradient (Fig. 2i) raises the possibility that proximal presubiculum is used mostly for planar 2D orientation, whereas the distal presubiculum may be involved in orientation in 3D.

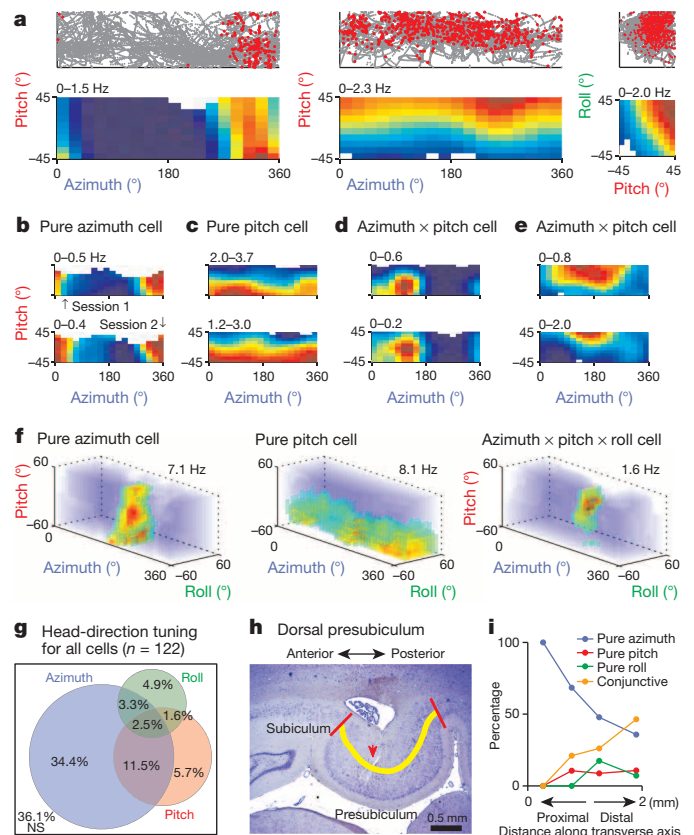


Figure 2 | Single cells encode one, two, or three Euler angles of head direction, and are organized according to a functional–anatomical gradient.

a, Example cells. Raw data (top; grey line, head direction; red dots, spikes), and corresponding 2D colour-coded rate-maps (bottom), showing the directional field for 3 neurons in 2D projections (pitch \times azimuth and roll \times pitch). The range of pitch and roll in the 2D plots was limited to match the behavioural range of angles covered by the bats (Fig. 1d). **b–e**, Examples of 4 neurons with different tuning properties, showing the 2D rate-maps for the two sessions (rows); minimal and maximal firing rates (in Hz) are indicated. **f**, Firing-rate maps in 3D projections (azimuth \times pitch \times roll) for a pure azimuth cell (left), pure pitch cell (middle), and triple-conjunctive cell tuned to all 3 Euler angles (right). **g**, Proportions of functional cell types in bat presubiculum (Methods). **h**, **i**, Functional–anatomical gradient of head-direction cells in dorsal presubiculum. **h**, Sagittal section through presubiculum; red lines, presubiculum borders. Yellow line: axis used to determine the anterior–posterior distance of the tetrode-track (arrowhead) from the subiculum border. **i**, Percentages of cells with pure tuning to azimuth, pitch, or roll, and conjunctive neurons tuned to any angular combination, plotted versus the recording position along the presubiculum’s transverse axis (Methods; graph based on all significant head-direction cells, $n = 78$; the four positional bins contain 8, 19, 23, 28 cells).

Azimuth tuning in inverted bats shifts by 180°

We next asked, what coordinate system is used by the brain to represent 3D head direction. The direction of a vector in 3D space is defined by azimuth and pitch, whereas roll is merely a rotation around this axis. As bats were shown to manoeuvre mostly in azimuth and pitch and much less in roll (ref. 46; see also Supplementary Video 1), and because conjunctive tuning to roll was relatively rare (Fig. 2g), we focused on azimuth and pitch. One way to dissociate between possible coordinate frames for representing azimuth \times pitch is to invert the animal, which is also ethologically relevant for bats.

We therefore introduced an inverted session (flanked between the two upright sessions), and found that nearly half of the azimuth-selective cells retained a directional firing when the bat was passively moved upside-down (Fig. 3a, middle row, blue lines; and Extended Data Fig. 5). Surprisingly, the cells' best tuning-angle in azimuth was shifted by 180°, relative to the best tuning angle in upright sessions 1 and 2 (Fig. 3a, middle row, magenta lines show 180°-shifted tuning curves; compare to the top and bottom rows). The 180°-shift was evident also at the population level: the differences in preferred azimuth between the two upright sessions were distributed around zero, indicating stability (Fig. 3b, top), whereas the differences between the upright and inverted sessions had a U-shaped distribution, peaking at $\pm 180^\circ$ (Fig. 3b, bottom and Extended Data Fig. 5d). Notably, the 180°-shift in the preferred azimuth of the neurons in inverted bats, was not caused by the fact that the inverted bat was moved passively (because passive-upright and active-upright tuning was very similar: see Extended Data Fig. 2a, b, f). The 180°-shift was also observed when bats had positioned themselves upside-down of their own volition (Extended Data Fig. 6). Interestingly, these results are different from findings in inverted rats, where many head-direction cells were reported to lose their directional properties^{33,34}.

Toroidal model of 3D head-direction cells

If head-direction cells simply encode the bearing of a vector in 3D space (which can be described in spherical coordinates, Fig. 3c), the neurons are expected to fire whenever the bat's head is directed towards the same distal point, irrespective of whether the bat is upright or inverted (Extended Data Fig. 7a, b). Our results of 180°-shift in the preferred direction under inversion, suggest that head direction in the bat pre-subiculum is not represented in spherical coordinates. Instead, it could be modelled by a toroidal manifold consisting of two independent cyclical degrees of freedom, 360° azimuth and 360° pitch (Fig. 3d, and Extended Data Fig. 7c, d). According to this model, pitching the head beyond $\pm 90^\circ$, does not change the azimuth angle in toroidal coordinates (Fig. 3d, red ring), and therefore a neuron tuned to a certain azimuth should keep firing also in the inverted position when the bat faces the 180°-opposite direction, as indeed we observed (Fig. 3a, b). Comparison of the correlations between upright and inverted sessions, computed in spherical versus toroidal coordinates, showed significantly higher correlations for the toroidal representation (Fig. 3e; t -test: $P < 0.001$; Methods).

Further, the toroidal representation entails that pure azimuth cells should appear as a vertical band on the torus (Fig. 3f, left column); pure pitch cells should form a horizontal band (Fig. 3f, middle); whereas conjunctive azimuth \times pitch cells should appear as a patch (Fig. 3f, right). This was indeed found for all three cell types (Fig. 3f, and Extended Data Figs 7 and 8).

Pitch cells exhibit circular tuning to 360°

The toroidal model implies that pitch is represented allocentrically and continuously over a circular range of 360° (Fig. 3d and Extended Data Fig. 9). To directly test this prediction, we recorded 175 cells from the pre-subiculum of 2 additional bats that crawled inside a vertical ring (Fig. 4a, pose 2). Pitch cells exhibited clear tuning (Fig. 4b, red curves, and Extended Data Fig. 10b, c), and the preferred pitch angles for different neurons were distributed across the entire 360° circular range of pitch, with a slight over-representation of 0° pitch (Fig. 4c; χ^2 test for

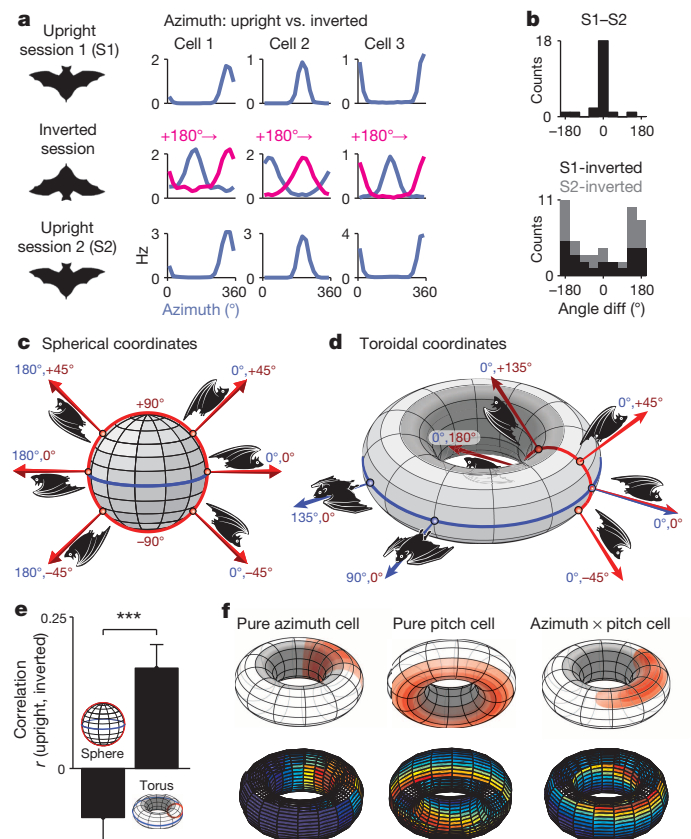


Figure 3 | Azimuth tuning in inverted bats is consistent with a toroidal model of head-direction coding. **a**, Azimuth tuning in three example cells, across the two upright behavioural sessions (top and bottom rows) and the inverted session (middle row, blue). Magenta curves in middle row: tuning-curves of the inverted session, shifted by 180°. **b**, Distributions of differences in preferred direction between sessions, for cells significantly tuned to azimuth in the inverted session. Top, angular difference between the two upright behavioural sessions (S1–S2), showing a peak in the distribution near 0°, indicating stability. Bottom, angular difference between each of the upright sessions and the inverted session (S1-inverted, S2-inverted), indicating 180°-shift of the preferred direction in the inverted session. **c**, Schematic depiction of azimuth (blue) and pitch (red) in spherical coordinates. Numerical values of azimuth and pitch are indicated with corresponding colours. Pitching beyond $\pm 90^\circ$ (along the red circle) results in abrupt shift of head azimuth by 180°. **d**, Toroidal coordinates, in which for each azimuth (blue) there is a 360° range of possible pitch-angles (red). Outer surface of the torus corresponds to upright bat positions; inner surface, to inverted positions. All positions along the red circle (corresponding to pitch angles between $\pm 180^\circ$) keep the same azimuth in the toroidal coordinates, hence pitching beyond $\pm 90^\circ$ does not change the azimuth (contrary to spherical coordinates, red circle in **c**). See also Extended Data Fig. 7a–d. **e**, Average 2D-correlations of firing maps in upright versus inverted sessions, for all cells tuned to azimuth or pitch ($n = 144$ (72 cells \times 2 sessions)), represented in either spherical (left) or toroidal coordinates (right). Error bars, mean \pm s.e.m.; *** $P < 0.001$. **f**, Toroidal 2D head-direction fields for a pure azimuth cell (left), pure pitch cell (middle), and conjunctive azimuth \times pitch cell (right). Upper panels, schematic illustrations. Lower panels, 3 example neurons; transparent bins, angles where the animal spent < 0.5 s.

non-uniformity: $P < 0.01$). The ‘ramp-like’ tuning often observed on the arena floor (Fig. 4b, black, corresponding to Fig. 4a, ‘pose 1’) was in fact a clipped version of the full circular tuning on the ring (Fig. 4b, compare black to red curves; population data, Fig. 4d). Similar to azimuth cells, individual pitch cells exhibited unimodal tuning (Fig. 4b, e, f and Extended Data Fig. 10b, c). Tuning widths for pitch cells were quite similar to those of azimuth cells (Fig. 4g and Extended Data Fig. 10). We thus conclude that the tuning to pitch was circular (360°), unimodal, and continuous.

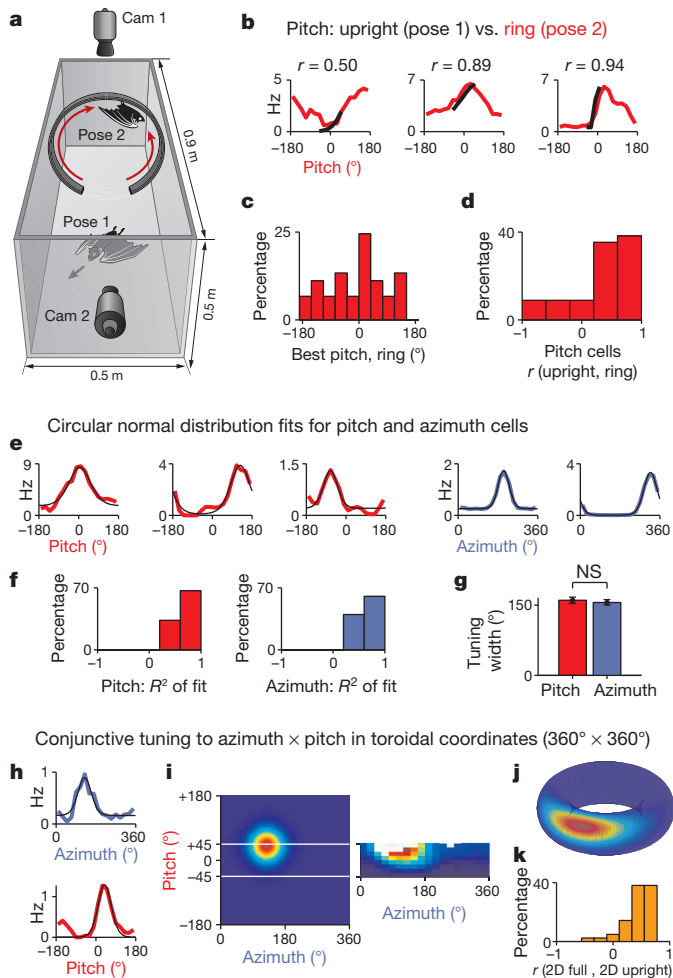


Figure 4 | Pitch cells are tuned unimodally and continuously to a cyclical range of 360°, as predicted by the toroidal model. **a**, Behavioural setup number 2 allowed a comparison of head-direction tuning when the bat crawled on the arena floor ('pose 1') versus crawling along the inner side of a vertically positioned ring ('pose 2'). **b**, Pitch tuning of 3 example cells during ring locomotion (red curve, pose 2, ranging between $\pm 180^\circ$), overlaid on the pitch tuning of the same neurons during floor locomotion (black curve, pose 1, ranging between $\pm 45^\circ$). **c**, Distribution of preferred firing directions of pitch-significant cells on the ring ($n = 45$); note that the population spans the full range of 360° . **d**, Distribution of correlation-coefficients for pitch tuning on ring versus floor. **e**, Examples of pitch and azimuth cells (red and blue curves, respectively), fitted with von Mises circular normal functions (black curves). **f**, Distributions of R^2 values of von Mises fits to pitch and azimuth. **g**, Average tuning width was similar for azimuth cells and pitch cells. Error bars, mean \pm s.e.m. **h**, Example of a conjunctive cell tuned to azimuth (top) and pitch (bottom), overlaid with von Mises fits (black curves). **i**, Left, full $360^\circ \times 360^\circ$ fitted 2D rate-map for the neuron in **h**, computed by multiplying the 1D von Mises fits to azimuth and pitch. Right, actual 2D rate map measured for this neuron during floor locomotion (the pitch range sampled on the floor is indicated by white lines on the left panel). **j**, Same $360^\circ \times 360^\circ$ fitted map as in **i**, presented in toroidal coordinates by wrapping the full 2D directional map around both axes. **k**, Distribution of correlation coefficients between the 2D map measured on the floor (as in **i**, right panel) and the relevant range of the full 2D map (as in **i**, left panel, the segment between the white lines), plotted for all the pitch and azimuth cells with sufficient number of spikes on both floor and ring ($n = 42$; Methods). NS, not significant.

Further, reconstruction of the full $360^\circ \times 360^\circ$ unwrapped maps for conjunctive azimuth \times pitch neurons, matched closely the actual receptive fields observed on the floor of the arena (Fig. 4h–k), as predicted by the toroidal model. Taken together, these results suggest a torus-based representation of 3D head direction in the bat presubiculum.

Head-direction cells in freely flying bats

An intriguing question is whether the 3D head-direction tuning that we observed in crawling bats would also be found during flight, where the movement is volumetric and not anchored to a specific locomotion plane. To test this, we used a wireless device to record 67 cells from the dorsal presubiculum during flight (Fig. 5a). We found that 30% of the cells ($n = 20/67$) were significantly directionally tuned, a very similar proportion to the 29% (78/266) found in crawling bats. Moreover, similar to crawling bats, the flight data revealed pure azimuth, pure pitch, and conjunctive azimuth \times pitch neurons (Fig. 5b). Tuning widths were also similar between flight and crawling (Fig. 5c). Interestingly, peak firing rates were more than twofold higher in flight as compared to crawling (Fig. 5d), consistent with the increased linear and angular velocity during flight³⁶. Collectively, these data suggest the existence of 3D head-direction tuning in both crawling and flight.

Discussion

Here we present the first evidence, to our knowledge, for a conjunctive representation of 3D head direction in the mammalian brain (Figs 1 and 2), which may serve as a '3D neural compass'. Head-direction cells were organized along the transverse axes of the presubiculum, with pure azimuth ('2D') cells being more prevalent in its proximal part and conjunctive 3D cells more abundant distally (Fig. 2h, i). Tuning widths for azimuth and pitch were rather similar (Fig. 4g). Notably, we found a surprising 180° -shift of the preferred-azimuth direction in inverted bats (Fig. 3), and a cyclical tuning of pitch cells to 360° -pitch (Fig. 4). These results, together with our finding of conjunctive cells, suggested that head direction is represented in toroidal coordinates (azimuth \times pitch).

The 3D head-direction cells described here, together with the 3D place-cells that we recently found in the bat hippocampus³⁶, could support 3D 'map-and-compass' navigation. We predict that conjunctive 3D head-direction cells might be found also in non-flying mammals that move in complex 3D environments or that orient their head up/down, such as squirrels, cats, dolphins and primates. Of particular interest will be to record from the presubiculum of mice which, in contrast to rats, naturally hang and locomote upside-down very well. We expect

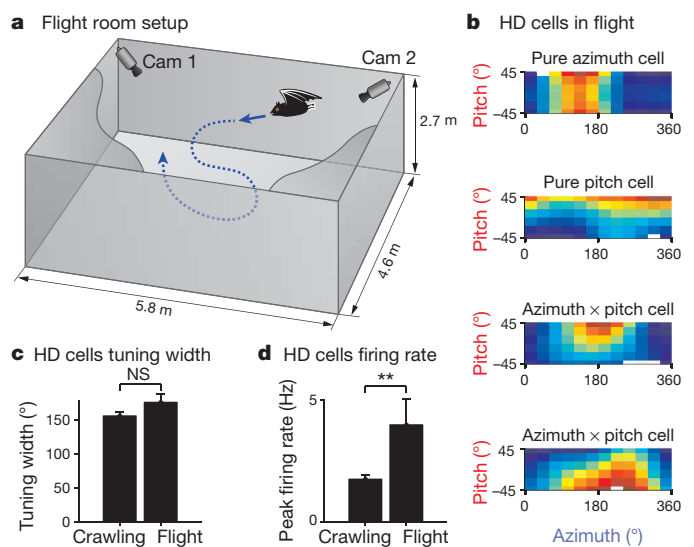


Figure 5 | Head-direction cells in freely flying bats. **a**, Behavioural setup number 3: flight-room. Objects and cues located in the room were omitted from the schematic, for clarity. **b**, Examples showing 2D rate-maps of head-direction cells recorded in flight, exhibiting different tuning properties: a pure azimuth cell, a pure pitch cell, and two azimuth \times pitch cells. **c**, Average tuning width for azimuth cells recorded during crawling ($n = 63$) and in flight ($n = 14$). Error bars, mean \pm s.e.m. **d**, Average peak firing rate for all the significant head-direction cells in crawling ($n = 78$) versus flight ($n = 20$). Firing rates increased more than twofold during flight. Error bars, mean \pm s.e.m.; ** $P < 0.01$. HD, head direction.

that, under inversion, head-direction cells in mice will exhibit similar properties to those in bats.

Our results yield interesting predictions for 3D grid cells. Proximal presubiculum (containing 2D head-direction cells, Fig. 2h, i) is known to project to distal medial entorhinal cortex (MEC), whereas distal presubiculum (containing 3D head-direction cells) projects to proximal MEC⁴⁷. Because computational models propose that head-direction information is used for path-integration to form grid cells^{43,48}, we might expect a gradient of 2D-to-3D grid cells in flying bats—that is, 2D grid cells (vertical columns) in distal MEC, near the border with lateral entorhinal cortex; and 3D grids (for example, 3D close-packing of spheres) in proximal MEC, near the border with parasubiculum.

The proposed toroidal model of head-direction coding is a generalization of the hemi-torus model for rat head-direction cells³², extended to the inverted pose. A fully toroidal topology has several advantages. First, it avoids undesirable network instabilities. In the spherical coordinate system, rotation of the head beyond $\pm 90^\circ$ pitch will result in an abrupt shift of the azimuth by 180° , and hence in abrupt changes of ensemble neural activity. In contrast, the toroidal topology confers continuity and stability, thus avoiding singularities in network activity when the bat passes through $\pm 90^\circ$ of pitch (see Supplementary Video 1 for example passages through $\pm 90^\circ$ of pitch). Second, an allocentric toroidal representation is commutative. Finally, a toroidal topology could directly support toroidal continuous attractor network models of head direction^{48–50}.

The torus is not necessarily the only possible model that can explain our data. The full set of rotations in 3D is described by ‘real projective space 3’ (RP³, also known as ‘hypersphere of rotations’), which has three degrees of freedom (azimuth, pitch, and roll), whereas the torus is a manifold within RP³, with only two degrees of freedom (azimuth \times pitch). However, for animals like bats—which exhibit very little roll, but manoeuvre substantially in azimuth and pitch—the torus is the simplest low-dimensional manifold within RP³ that is consistent with our data.

Finally, we note that head-azimuth was defined to date as the direction of the animal’s nose (rostral–caudal axis) in the horizontal plane; however, the toroidal representation can be naturally interpreted as a coordinate system with two orthogonal axes: the azimuthal direction is anchored to the interaural axis (the animal’s ears), and thus remains the same for any pitch, whereas the pitch direction is being anchored to the rostral–caudal axis.

In summary, the toroidal topology of 3D head-direction representation in bats is consistent with the animal’s natural behaviour and the need to represent absolute 3D directions in space, as well as with biological requirements for network stability, and with the structure of continuous attractor models. Thus, it provides a convergence of ethology, neurobiology and computation.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 25 July; accepted 4 November 2014.

Published online 3 December 2014.

- O’Keefe, J. & Nadel, L. *The Hippocampus as a Cognitive Map* (Oxford Univ. Press, 1978).
- Gallistel, C. R. *The Organization of Learning* (MIT Press, 1990).
- Menzel, R. et al. Honey bees navigate according to a map-like spatial memory. *Proc. Natl Acad. Sci. USA* **102**, 3040–3045 (2005).
- Taube, J. S. The head direction signal: origins and sensory-motor integration. *Annu. Rev. Neurosci.* **30**, 181–207 (2007).
- Wu, L. Q. & Dickman, J. D. Neural correlates of a magnetic sense. *Science* **336**, 1054–1057 (2012).
- O’Keefe, J. & Dostrovsky, J. The hippocampus as a spatial map. Preliminary evidence from unit activity in the freely-moving rat. *Brain Res.* **34**, 171–175 (1971).
- Wilson, M. A. & McNaughton, B. L. Dynamics of the hippocampal ensemble code for space. *Science* **261**, 1055–1058 (1993).
- Leutgeb, S. et al. Independent codes for spatial and episodic memory in hippocampal neuronal ensembles. *Science* **309**, 619–623 (2005).
- Ulanovsky, N. & Moss, C. F. Hippocampal cellular and network activity in freely moving echolocating bats. *Nature Neurosci.* **10**, 224–233 (2007).
- Harvey, C. D., Collman, F., Dombeck, D. A. & Tank, D. W. Intracellular dynamics of hippocampal place cells during virtual navigation. *Nature* **461**, 941–946 (2009).
- Royer, S. et al. Control of timing, rate and bursts of hippocampal place cells by dendritic and somatic inhibition. *Nature Neurosci.* **15**, 769–775 (2012).
- Miller, J. F. et al. Neural activity in human hippocampal formation reveals the spatial context of retrieved memories. *Science* **342**, 1111–1114 (2013).
- Hafting, T., Fyhn, M., Molden, S., Moser, M.-B. & Moser, E. I. Microstructure of a spatial map in the entorhinal cortex. *Nature* **436**, 801–806 (2005).
- Barry, C., Hayman, R., Burgess, N. & Jeffery, K. J. Experience-dependent rescaling of entorhinal grids. *Nature Neurosci.* **10**, 682–684 (2007).
- Boccarra, C. N. et al. Grid cells in pre- and parasubiculum. *Nature Neurosci.* **13**, 987–994 (2010).
- Yartsev, M. M., Witter, M. P. & Ulanovsky, N. Grid cells without theta oscillations in the entorhinal cortex of bats. *Nature* **479**, 103–107 (2011).
- Jacobs, J. et al. Direct recordings of grid-like neuronal activity in human spatial navigation. *Nature Neurosci.* **16**, 1188–1190 (2013).
- Solstad, T., Boccarra, C. N., Kropff, E., Moser, M.-B. & Moser, E. I. Representation of geometric borders in the entorhinal cortex. *Science* **322**, 1865–1868 (2008).
- Savelli, F., Yoganarasimha, D. & Knierim, J. J. Influence of boundary removal on the spatial representations of the medial entorhinal cortex. *Hippocampus* **18**, 1270–1282 (2008).
- Lever, C., Burton, S., Jeewajee, A., O’Keefe, J. & Burgess, N. Boundary vector cells in the subiculum of the hippocampal formation. *J. Neurosci.* **29**, 9771–9777 (2009).
- Taube, J. S., Muller, R. U. & Ranck, J. B. Jr. Head-direction cells recorded from the postsubiculum in freely moving rats. I. Description and quantitative analysis. *J. Neurosci.* **10**, 420–435 (1990).
- Taube, J. S., Muller, R. U. & Ranck, J. B. Jr. Head-direction cells recorded from the postsubiculum in freely moving rats. II. Effects of environmental manipulations. *J. Neurosci.* **10**, 436–447 (1990).
- Zugaro, M. B., Berthoz, A. & Wiener, S. I. Background, but not foreground, spatial cues are taken as references for head direction responses by rat anterodorsal thalamus neurons. *J. Neurosci.* **21**, RC154 (2001).
- Sargolini, F. et al. Conjunctive representation of position, direction, and velocity in entorhinal cortex. *Science* **312**, 758–762 (2006).
- Langston, R. F. et al. Development of the spatial representation system in the rat. *Science* **328**, 1576–1580 (2010).
- Wills, T. J., Cacucci, F., Burgess, N. & O’Keefe, J. Development of the hippocampal cognitive map in preweanling rats. *Science* **328**, 1573–1576 (2010).
- Valerio, S. & Taube, J. S. Path integration: how the head direction signal maintains and corrects spatial orientation. *Nature Neurosci.* **15**, 1445–1453 (2012).
- Brandon, M. P., Bogaard, A. R., Schultheiss, N. W. & Hasselmo, M. E. Segregation of cortical head direction cell assemblies on alternating theta cycles. *Nature Neurosci.* **16**, 739–748 (2013).
- Wallace, D. J. et al. Rats maintain an overhead binocular field at the expense of constant fusion. *Nature* **498**, 65–69 (2013).
- Yovel, Y., Falk, B., Moss, C. F. & Ulanovsky, N. Optimal localization by pointing off axis. *Science* **327**, 701–704 (2010).
- Knierim, J. J., McNaughton, B. L. & Poe, G. R. Three-dimensional spatial selectivity of hippocampal neurons during space flight. *Nature Neurosci.* **3**, 209–210 (2000).
- Stackman, R. W., Tullman, M. L. & Taube, J. S. Maintenance of rat head direction cell firing during locomotion in the vertical plane. *J. Neurophysiol.* **83**, 393–405 (2000).
- Taube, J. S., Stackman, R. W., Calton, J. L. & Oman, C. M. Rat head direction cell responses in zero-gravity parabolic flight. *J. Neurophysiol.* **92**, 2887–2997 (2004).
- Calton, J. L. & Taube, J. S. Degradation of head direction cell activity during inverted locomotion. *J. Neurosci.* **25**, 2420–2428 (2005).
- Hayman, R., Verriotti, M. A., Jovalekic, A., Fenton, A. A. & Jeffery, K. J. Anisotropic encoding of three-dimensional space by place cells and grid cells. *Nature Neurosci.* **14**, 1182–1188 (2011).
- Yartsev, M. M. & Ulanovsky, N. Representation of three-dimensional space in the hippocampus of flying bats. *Science* **340**, 367–372 (2013).
- Rubin, A., Yartsev, M. M. & Ulanovsky, N. Encoding of head direction by hippocampal place cells in bats. *J. Neurosci.* **34**, 1067–1080 (2014).
- Taube, J. S., Wang, S. S., Kim, S. Y. & Frohardt, R. J. Updating of the spatial reference frame of head direction cells in response to locomotion in the vertical plane. *J. Neurophysiol.* **109**, 873–888 (2013).
- Stackman, R. W. & Taube, J. S. Firing properties of rat lateral mammillary single units: head direction, head pitch, and angular head velocity. *J. Neurosci.* **18**, 9020–9037 (1998).
- Bassett, J. P. & Taube, J. S. Neural correlates for angular head velocity in the rat dorsal tegmental nucleus. *J. Neurosci.* **21**, 5740–5751 (2001).
- Calton, J. L. et al. Hippocampal place cell instability after lesions of the head direction cell network. *J. Neurosci.* **23**, 9719–9731 (2003).
- Bonnevie, T. et al. Grid cells require excitatory drive from the hippocampus. *Nature Neurosci.* **16**, 309–317 (2013).
- Burak, Y. & Fiete, I. R. Accurate path integration in continuous attractor network models of grid cells. *PLoS Comput. Biol.* **5**, e1000291 (2009).
- Canto, C. B., Koganezawa, N., Beed, P., Moser, E. I. & Witter, M. P. All layers of medial entorhinal cortex receive presubicular and parasubicular inputs. *J. Neurosci.* **32**, 17620–17631 (2012).
- Jeffery, K. J., Jovalekic, A., Verriotti, M. & Hayman, R. Navigating in a three-dimensional world. *Behav. Brain Sci.* **36**, 523–543 (2013).
- Iriarte-Diaz, J. & Swartz, S. M. Kinematics of slow turn maneuvering in the fruit bat *Cynopterus brachyotis*. *J. Exp. Biol.* **211**, 3478–3489 (2008).

Human gut Bacteroidetes can utilize yeast mannan through a selfish mechanism

Fiona Cuskin^{1,2*}, Elisabeth C. Lowe^{1*}, Max J. Temple^{1*}, Yanping Zhu^{1,2}, Elizabeth A. Cameron³, Nicholas A. Pudlo³, Nathan T. Porter³, Karthik Urs³, Andrew J. Thompson⁴, Alan Cartmell⁵, Artur Rogowski¹, Brian S. Hamilton⁶, Rui Chen⁷, Thomas J. Tolbert⁸, Kathleen Piens⁹, Debby Bracke⁹, Wouter Vervecken⁹, Zalihe Hakki⁵, Gaetano Speciale⁵, Jose L. Munõz-Munõz¹, Andrew Day¹, Maria J. Peña², Richard McLean¹⁰, Michael D. Suits¹¹, Alisdair B. Boraston¹¹, Todd Atherly¹², Cherie J. Ziemer^{12‡}, Spencer J. Williams⁵, Gideon J. Davies⁴, D. Wade Abbott^{2,10}, Eric C. Martens³ & Harry J. Gilbert^{1,2}

Yeasts, which have been a component of the human diet for at least 7,000 years, possess an elaborate cell wall α -mannan. The influence of yeast mannan on the ecology of the human microbiota is unknown. Here we show that yeast α -mannan is a viable food source for the Gram-negative bacterium *Bacteroides thetaiotaomicron*, a dominant member of the microbiota. Detailed biochemical analysis and targeted gene disruption studies support a model whereby limited cleavage of α -mannan on the surface generates large oligosaccharides that are subsequently depolymerized to mannose by the action of periplasmic enzymes. Co-culturing studies showed that metabolism of yeast mannan by *B. thetaiotaomicron* presents a 'selfish' model for the catabolism of this difficult to breakdown polysaccharide. Genomic comparison with *B. thetaiotaomicron* in conjunction with cell culture studies show that a cohort of highly successful members of the microbiota has evolved to consume sterically-restricted yeast glycans, an adaptation that may reflect the incorporation of eukaryotic microorganisms into the human diet.

The microbial community in the human large bowel, the microbiota^{1,2}, is central to the health and nutrition of its host^{3–6}. Glycan utilization is a key evolutionary driver underpinning the structure^{1,2} of this microcosm^{1,2}, with the Bacteroidetes playing a dominant role in this process. The genomes of Bacteroidetes contain polysaccharide utilization loci (PULs)⁷ that encode the apparatus required to utilize complex carbohydrates, with each PUL orchestrating the degradation of a specific glycan.

The microbiota contains a cohort of bacteria that target α -mannosidic linkages (based on sequence similarity to glycoside hydrolase families (GHs) that are populated by α -mannosidases and α -mannanases)^{8,9}, indicating that α -mannose-containing glycans, such as those from yeast and other fungal α -mannans, are significant nutrients for these microbes (see Supplementary Information section 1.0). Furthermore, these glycans are implicated in the immunopathology of the inflammatory bowel disease, Crohn's disease^{10,11} (Supplementary Information section 2.0). The genome of *Bacteroides thetaiotaomicron*, a dominant member of the microbiota, encodes 36 proteins predicted to display α -mannosidase or α -mannanase activity¹². Here we unravel the mechanism by which *B. thetaiotaomicron* metabolizes the major α -mannose-containing glycans presented to the large bowel. The data also show that *B. thetaiotaomicron* expresses a specific yeast α -mannan degrading system that is distinct from the high mannose mammalian *N*-glycan (HMNG) depolymerizing apparatus.

Yeast α -mannan utilization by *B. thetaiotaomicron*

B. thetaiotaomicron utilizes α -mannan as a sole carbon source and transcriptional studies identified three *B. thetaiotaomicron* PULs (MAN-PUL1, MAN-PUL2 and MAN-PUL3; Fig. 1a) that were activated by α -mannan

from *Saccharomyces cerevisiae*¹³, *Schizosaccharomyces pombe* and the pathogenic yeast *Candida albicans* (Fig. 1b and Extended Data Fig. 1a, b). *B. thetaiotaomicron* mutants lacking MAN-PUL2 or MAN-PUL1/2/3 were unable to grow on yeast mannan *in vitro*, (Extended Data Fig. 1c). In gnotobiotic mice fed a diet lacking glycans the *B. thetaiotaomicron* mutant Δ MAN-PUL1/2/3 outcompeted the wild-type bacterium, whereas wild-type *B. thetaiotaomicron* was the dominant species in rodents fed a yeast-mannan-rich diet, (Fig. 2a). These data underscore the importance of MAN-PUL1/2/3 when *B. thetaiotaomicron* is exposed to yeast mannan *in vivo*, while the intriguing dominance of the mutant strain in animals fed a polysaccharide-free diet is further considered in Supplementary Information section 3.2. To explore whether *B. thetaiotaomicron* degraded α -mannan and HMNG¹³ by distinct enzyme systems, the PULs activated by a HMNG, Man₈GlcNAc₂, were evaluated. A single PUL was activated by Man₈GlcNAc₂, which was distinct from MAN-PUL1/2/3 (Fig. 1c), demonstrating that degradation, and thus utilization, of α -mannan and HMNG are carried out by different PULs.

Analysis of the growth profiles of 29 human gut Bacteroidetes species revealed that nine species metabolized *S. cerevisiae* α -mannan, with 33 out of 34 strains of *B. thetaiotaomicron* growing on the glycan (Fig. 1d). These data show that *B. thetaiotaomicron*, along with some of its phylogenetically related neighbours, dominates yeast mannan metabolism in the Bacteroidetes phylum of the microbiota. Within fully sequenced genomes of 177 microbiota members, the presence of GHs integral to α -mannan degradation (GH38, GH92 and GH76) are restricted to five species of *Bacteroides* and three of *Parabacteroides*; no Firmicutes contained this grouping of glycoside hydrolases⁸.

¹Institute for Cell and Molecular Biosciences, Newcastle University, Newcastle-upon-Tyne NE2 4HH, UK. ²Complex Carbohydrate Research Center, The University of Georgia, 315 Riverbend Road, Athens, Georgia 30602, USA. ³Department of Microbiology and Immunology, University of Michigan Medical School, Ann Arbor, Michigan 48109 USA. ⁴Department of Chemistry, University of York, York YO10 5DD, UK. ⁵School of Chemistry and Bio21 Molecular Science and Biotechnology Institute, University of Melbourne, Parkville, Victoria 3010, Australia. ⁶Interdisciplinary Biochemistry Graduate Program, Indiana University, 800 E. Kirkwood Avenue, Bloomington, Indiana 47405, USA. ⁷Department of Chemistry, Indiana University, 800 E. Kirkwood Avenue, Bloomington, Indiana 47405, USA. ⁸Department of Pharmaceutical Chemistry, University of Kansas School of Pharmacy, 2095 Constant Avenue, Lawrence, Kansas 66047, USA. ⁹Oxyrane, 9052 Ghent, Belgium. ¹⁰Agriculture and Agri-Food Canada, Lethbridge Research Centre, Lethbridge, Alberta T1J 4B1, Canada. ¹¹Biochemistry and Microbiology, University of Victoria, Victoria, British Columbia V8P 5C2, Canada. ¹²USDA, Agricultural Research Service, National Laboratory for Agriculture and the Environment, Ames, Iowa 50011, USA.

*These authors contributed equally to this work.

‡Deceased.

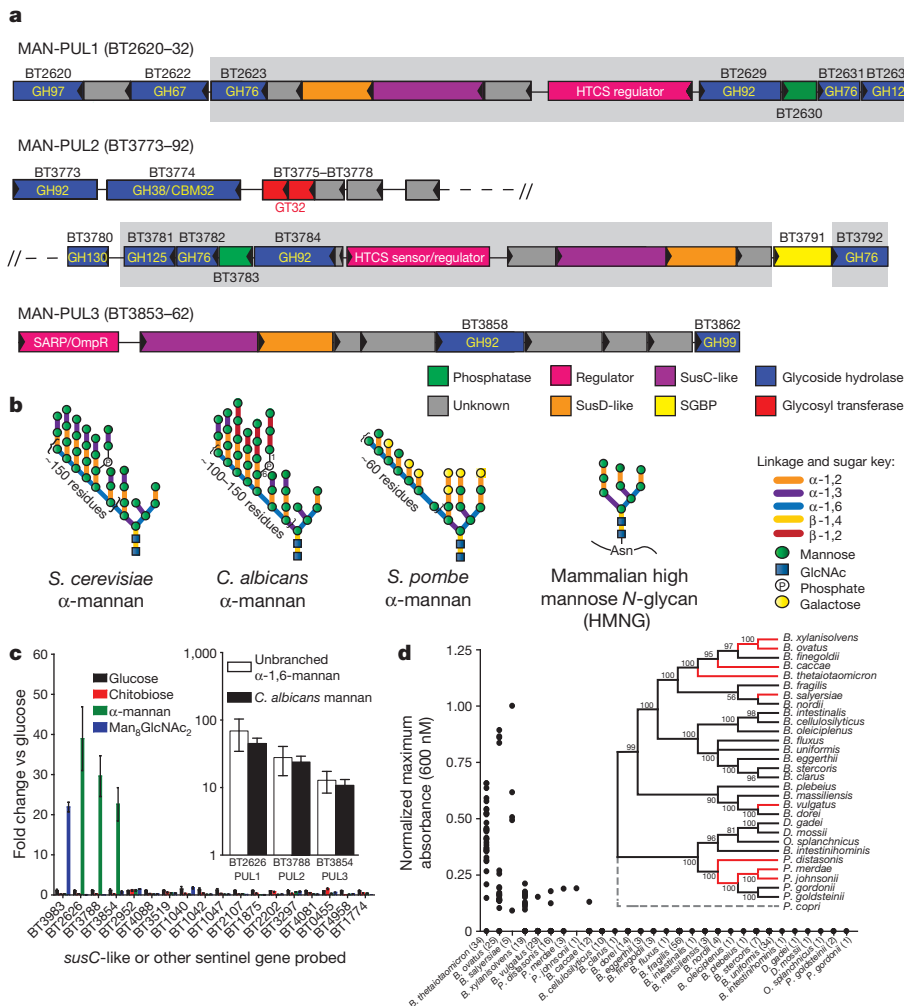


Figure 1 | *B. thetaiotaomicron* PULs involved in yeast α -mannan metabolism, and utilization of the glycan in Bacteroidetes. **a**, Genes encoding known or predicted functionalities are colour-coded and, where appropriate, annotated according to their CAZy family. SGBP; surface glycan binding protein. **b**, Structures of yeast (*S. cerevisiae*) mannan and HMNG. **c**, Cells were grown on media containing glycans as the sole carbon source ($n = 3$ separate cultures per substrate), and *susC* transcripts derived from PULs that encode GH76 and/or, GH92 enzymes were quantified by qPCR. The fold change relative to glucose-grown cells is shown on the y axis and the x axis shows the genes probed. Inset: levels of *susC* transcripts derived from MAN-PUL1/2/3 when *B. thetaiotaomicron* was cultured on mannan comprising only the α -1,6-Man backbone or *Candida albicans* mannan. Data are averages and standard deviations of three biological replicates. **d**, Strains of 29 different bacterial species of Bacteroidetes (number of strains in parentheses) were inoculated into minimal media containing *S. cerevisiae* mannan as the sole carbon source ($n = 2$ replicate cultures per strain). Growth measured at 48 h at A_{600nm} . Inset: phylogeny of Bacteroidetes species with bootstrap numbers indicated; organisms growing on mannan connected by red lines. The non-Bacteroides genera were *Odoribacter* (O), *Parabacteroides* (P) and *Dysgonomonas* (D).

Structure of the *B. thetaiotaomicron* α -mannan PULs

B. thetaiotaomicron MAN-PUL1 and MAN-PUL2 display significant synteny (Supplementary Table 1) while MAN-PUL3 displays no organizational similarity to the other two loci (Fig. 1a). Characterization of the 15 enzymes encoded by the mannan PULs revealed that these loci orchestrate α -mannan degradation from diverse yeasts and possibly other fungi. For example, MAN-PUL1 contains an α -galactosidase, BT2620, which targets α -galactosyl linkages absent in *S. cerevisiae* mannan but present in other fungal α -mannans such as the yeast *Schizosaccharomyces pombe*¹⁴ (Extended Data Fig. 1d), explaining why inactivation of MAN-PUL1 affects growth of *B. thetaiotaomicron* on this polysaccharide (Extended Data Fig. 1b). Functional diversity is also evident in MAN-PUL2, which, in addition to its catabolic role, encodes glycosyltransferases that mediate synthesis of the trisaccharide Man- α -1,3-[Man- α -1,6]-Man (Extended Data Fig. 2). Thus, MAN-PUL2 comprises a unique example of the co-regulation of related catabolic and biosynthetic functions within a single PUL¹³.

Mannan degradation at the cell surface and periplasm

The enzymatic degradation of α -mannan is restricted through steric constraints imposed through the side chains appended to the backbone (Supplementary Table 2 and Extended Data Figs 1e and 4a). Critically, the α -1,6-Man-backbone is not accessible to the endo- α -1,6-mannanases until the side chains have been removed, consistent with the topography of the substrate binding cleft of these enzymes (Extended Data Fig. 3a, b). The MAN-PUL3-encoded endo- α -1,2-mannosidase BT3862 released a limited proportion of the terminal Man- α -1,3-Man disaccharides, thereby assisting in exposing the backbone. The structural basis for the specificity displayed by BT3862 is

described in Extended Data Fig. 3c, d. In addition to BT3862, other surface enzymes are required to partially expose the α -1,6-Man-backbone. Screening for such enzymes revealed an α -mannosidase, BT2199, that mediates very limited extracellular removal of side chains (Extended Data Fig. 4a) and thus facilitates limited cleavage of the mannan backbone by the surface endo- α -1,6-mannanases BT2623 and BT3792. The activity profiles of BT2199 and the surface endo-acting mannanases are consistent with the observed production of large oligosaccharides by non-growing cells of *B. thetaiotaomicron* incubated with yeast mannan (Extended Data Fig. 4b, c).

The cellular location of the key α -mannan hydrolysing enzymes (Fig. 2b, c) indicates that the polysaccharide was degraded primarily in the periplasm where the side chains were removed by the synergistic action of α -mannosidases and sugar-6-monophosphatases¹⁵ (Fig. 2d, e, Extended Data Fig. 5, Supplementary Tables 2 and 3 and Supplementary Information section 4.0). The broad specificity of BT3774 enabled the α -mannosidase to play a key role in the removal of the uncharged side chains, being the only enzyme capable of removing the sterically restricted α -1,2-Man units linked to the α -mannan backbone (Fig. 3 and Extended Data Figs 4f and 5) as well as cleaving the Man-1-phosphate linkage, a critical step in the removal of the phosphorylated branches (Fig. 2d). The importance of BT3774 in α -mannan catabolism was consistent with the growth profile of the *B. thetaiotaomicron* mutant *Abt3774* (Extended Data Fig. 1d).

Hydrolysis of the α -1,6-Man backbone of α -mannan by *B. thetaiotaomicron* was mediated by surface and periplasmic 'retaining' endo- α -1,6-mannanases (Extended Data Fig. 4e). The surface endo-mannanases BT2623 and BT3792 (Fig. 2b, c) generated large oligosaccharides, while the periplasmic mannanases BT2631 and BT3782 produced small limit

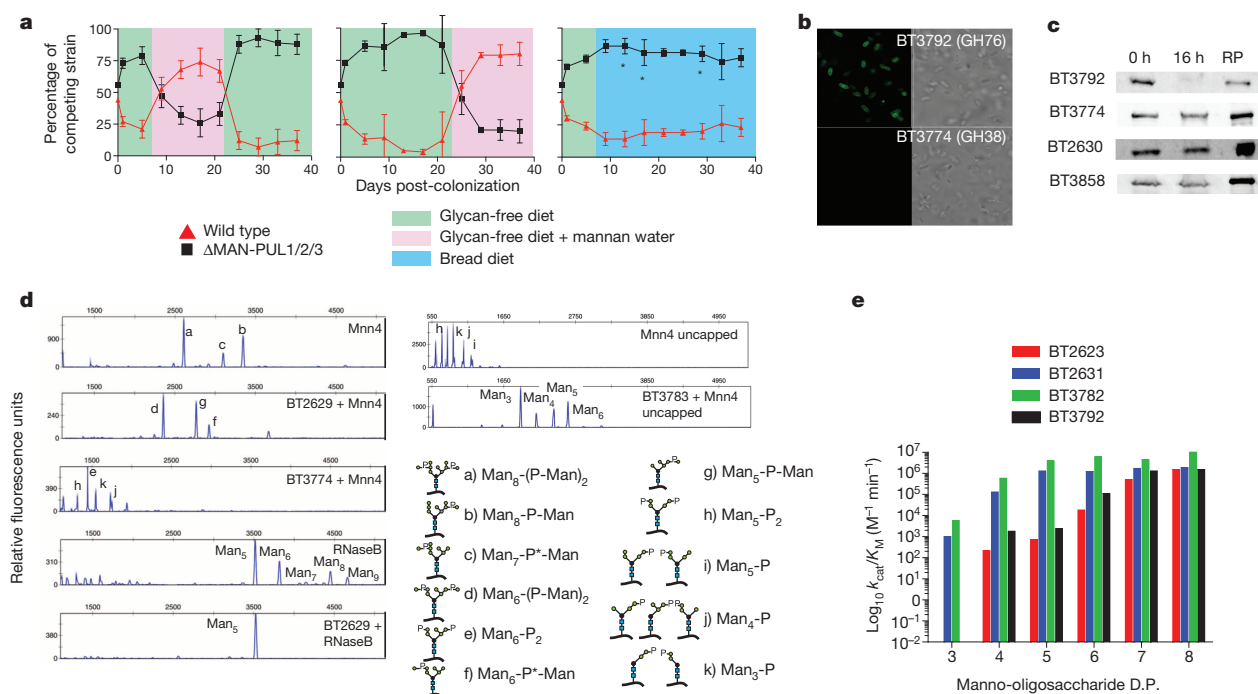


Figure 2 | Mannan PULs enable colonization of gnotobiotic mice; key biochemical and cellular features of the encoded enzymes. **a**, Colonization of gnotobiotic mice ($n = 5$) by wild type *B. thetaiotaomicron* (red) and a mutant (black) lacking MAN-PUL1/2/3 (Δ MAN-PUL1/2/3). On day 0 mice were gavaged with $\sim 10^8$ colony-forming units (c.f.u.) of 50:50 of the two *B. thetaiotaomicron* strains and then fed a control diet lacking *B. thetaiotaomicron*-digestible glycans (green). Left, after 7 days the control diet was supplemented with 1% YM in drinking water (pink shading). Middle, identical treatment as to the left panel except no mannan was included after day 7 (shaded green). At day 21 mannan in the water was switched between the groups in the left and middle panels, indicated by the colour panels (pink = mannan, green = no mannan). Right, mice fed the control diet (shaded green) were switched to a diet containing leavened bread (blue). The average abundance of the mannan utilization mutant (black) in mice fed the bread diet compared to the corresponding time points in mice fed the glycan-free diet (mean \pm s.e.m., $81\% \pm 1.2\%$ in animals fed the bread diet versus $90\% \pm 1.7\%$ in mice fed the glycan-free diet; $P = 0.00005$ by unpaired Student's t -test). Time points at which there was a significant difference between the mutant and wild

type *B. thetaiotaomicron* on the bread diet compared to the glycan-free diet ($P \leq 0.05$) are indicated with an asterisk. **b**, Fluorescent- and light-microscopy images of *B. thetaiotaomicron* incubated with polyclonal antibodies against BT3792 and BT3774. **c**, Western blots of *B. thetaiotaomicron* cells cultured on yeast mannan that were untreated with proteinase K (0 h) or incubated with 2 mg ml^{-1} proteinase K for 16 h. RP, recombinant protein. Blots were probed with antibodies against the *B. thetaiotaomicron* enzymes. Localization blots and microscopy images (**b**, **c**) are representative data from three biological replicates. **d**, Phosphorylated high mannose *N*-glycan MNN4 and RNAase B incubated with BT2629 (GH92 α -mannosidase), BT3783 (phosphatase) or BT3774 (GH38 α -mannosidase) and the products analysed by capillary electrophoresis. **e**, Catalytic efficiency of the endo- α -1,6-mannanases against α -1,6-mannooligosaccharides (see Supplementary Information Table 4 for full data). D.P., degree of polymerization of the mannooligosaccharides.

products (Extended Data Fig. 6a–c). Consistent with these product profiles, the periplasmic enzymes were ~ 100 - to $1,000$ -fold more active than the surface mannanases against oligosaccharides with a degree of polymerization (d.p.) < 6 , whereas against large oligosaccharides (d.p. ≥ 6) the periplasmic and surface enzymes displayed broadly similar activities (Fig. 2f and Supplementary Table 4). The linear and hence unadorned α -1,6-mannooligosaccharides generated in the periplasm were hydrolyzed by the exo- α -1,6-mannosidases BT2632 and BT3781 (Extended Data Fig. 6d, e and Supplementary Table 2), as previously proposed¹⁶. The inability of the backbone-cleaving enzymes to attack wild-type yeast α -mannan indicates that these glycoside hydrolases cannot accommodate α -1,6-linked Man decorated at O2 (side chains are appended α -1,2 to the mannan backbone) in the active site or proximal subsites, consistent with the structures of the substrate binding clefts of BT3792 and BT3781 (Extended Data Fig. 3a–f).

A *B. thetaiotaomicron* PUL for HMNG catabolism

HMNG-PUL encodes four enzymes and two surface glycan binding proteins (Extended Data Fig. 7a). BT3990 and BT3991 target α -1,2-Man and α -1,3-Man linkages, respectively, in HMNGs¹⁷. The terminal undecorated α -1,6-Man exposed by BT3990 is hydrolysed by BT3994, which requires GlcNAc at the reducing end for activity, producing Man- α -1,6-Man- β 1,4-GlcNAc (Extended Data Fig. 7b). BT3987 is a surface

endo-acting *N*-acetylglucosaminidase, which cleaves the oligosaccharide from its polypeptide (Extended Data Fig. 7c, d). The released HMNG is held on the surface of *B. thetaiotaomicron* through the mannose-binding protein BT3986, while the SusD homologue BT3984, by recognizing GlcNAc at the reducing end of the glycan (Extended Data Fig. 7e, f and Supplementary Table 5), probably orientates the glycan into the outer membrane porin (SusC homologue BT3983) for transport into the periplasm, where the periplasmic α -mannosidases hydrolyse the oligosaccharide into a trisaccharide that is degraded by enzymes that are not encoded by HMNG-PUL. A model for the enzymatic degradation of HMNG is displayed in Fig. 4. The importance of HMNG-PUL in the metabolism of the *N*-glycan is consistent with the reduced growth of the *Abt3993* mutant (lacking the extra-cellular σ factor regulator of HMNG-PUL) cultured on Man₈GlcNAc₂ (Extended Data Fig. 7g).

Yeast α -mannan degradation by *B. thetaiotaomicron*

Substrate promiscuity is a hallmark of the α -mannan degradation apparatus of *B. thetaiotaomicron*, as steric constraints limit enzymatic access to the glycan side chains, and consequently the polysaccharide backbone. Accordingly, *B. thetaiotaomicron* uniquely uses surface α -mannosidase(s), probably BT2199, to generate limited, but sufficient, side chain removal for the surface endo-acting enzymes to mediate infrequent cleavage of the backbone. The *B. thetaiotaomicron* mannan-degrading apparatus

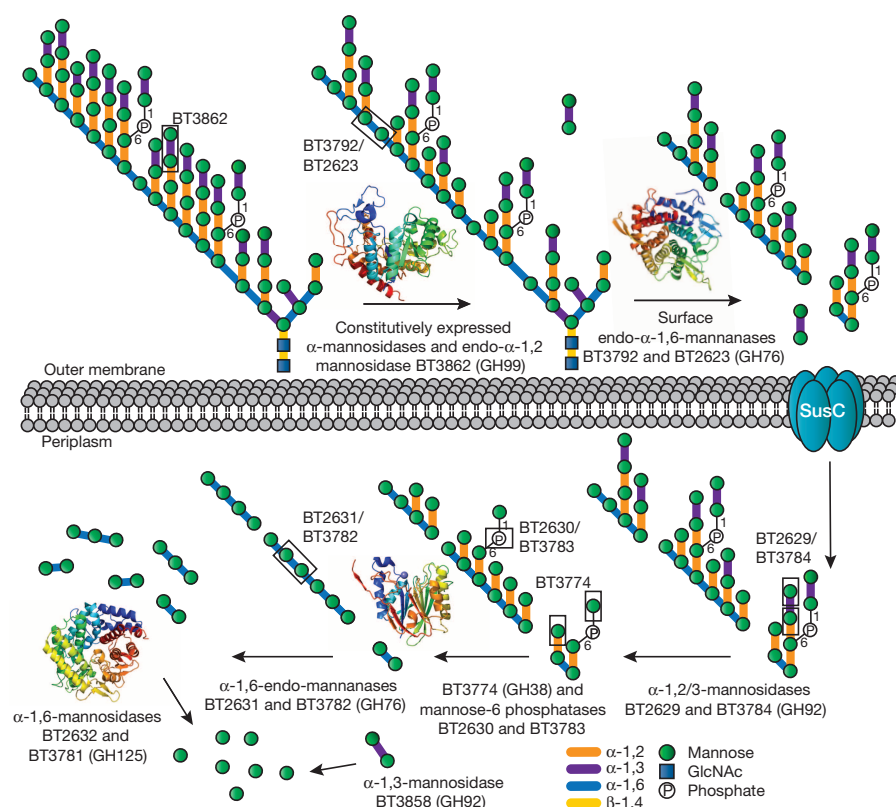


Figure 3 | Model of yeast mannan deconstruction by *B. thetaiotaomicron*. Boxes show examples of bonds cleaved by the endo- α -1,2-mannosidase BT3862, by the α -1,3- and α -1,2-mannosidase activities displayed by BT2629 and BT3784, respectively, and the Man-1-P and α -1,2-Man linkages targeted by BT3774. Extended Data Fig. 3 shows 3D structures (cartoons, colour-ramped from the amino terminus (blue) to the carboxy terminus (red)) of the enzymes that play a key role in mannan degradation. In this model limited degradation occurs at the surface and the bulk of glycan degradation occurs in the periplasm. SusC-like proteins mediate transport across the outer membrane⁷.

is optimized to minimize nutrient loss, illustrated by the observation that no oligosaccharides were released into the culture supernatant when *B. thetaiotaomicron* was cultured on α -mannan (Extended Data Fig. 4c).

Endo- α -1,6-mannanases are central to the hierarchical degradation of yeast mannan, reflecting their location and substrate specificities. The surface endo-acting enzymes generate large oligosaccharides minimizing extracellular metabolism and loss to other microbiota residents. The periplasmic endo- α -1,6-mannanases generate numerous short oligosaccharides, maximizing the substrate available to the periplasmic exo-acting α -1,6-mannosidases. This 'selfish' model is consistent with the inability of *B. thetaiotaomicron* to support growth of mannose- and

mannan-backbone-using strains of *Bacteroides* on intact *S. cerevisiae* α -mannan (Fig. 5). It should be emphasized, however, that the degradation of at least some polysaccharides is mediated through synergistic interactions between different members of the microbiota^{18,19}, illustrating the diverse mechanisms by which nutrients are used in this microbial community.

This work provides insights into the adaptation of the microbiota to yeast domestication in the human diet reflecting the regular consumption of yeast-leavened bread, fermented beverages and products such as soy sauce. *B. thetaiotaomicron*, and a limited number of other microbiota-derived Bacteroidetes, have evolved a complex machinery to digest and metabolize yeast cell-wall mannans. Phylogenetic analysis suggests that this trait penetrated the human gut Bacteroidetes at least twice, once each in the *Bacteroides* and *Parabacteroides*. It is also possible, however, that close relatives of *B. thetaiotaomicron* have gained parts of this complex catabolic trait via separate events. Analysis of 250 human metagenomic samples revealed α -mannan-degrading PULs closely related to those from *B. thetaiotaomicron* in a majority (62%) of subjects (Extended Data Fig. 8) making it more common than culturally-restricted traits like red algal porphyran degradation²⁰, but less common than plant cell-wall xyloglucan degrading PULs, identified in divergent *Bacteroides* species^{21,22} (see Supplementary Information section 3.3).

The α -mannan degrading capacity of *B. thetaiotaomicron* is consistent with the bacterium's 'glycan generalist' strategy^{7,12,23}, enabling the microorganism to thrive in the competitive environment of the microbiota, where the omnivorous diet of the host requires rapid adaptation to the nutrients presented to the distal gut. Many of the organisms in the microbiota produce enzymes that attack the major components of the human diet such as starch and pectins. An additional survival strategy for some members of the microbiota, such as *B. thetaiotaomicron*, is the targeting of low-abundance, highly complex dietary glycans that are not metabolized by most other organisms, exemplified here by yeast mannan. With respect to the human host the degradation of α -mannan by *B. thetaiotaomicron* may also be relevant to the health-promoting effects of the human microbiota²⁴. This work provides insights into a sophisticated α -mannan degrading apparatus that exists

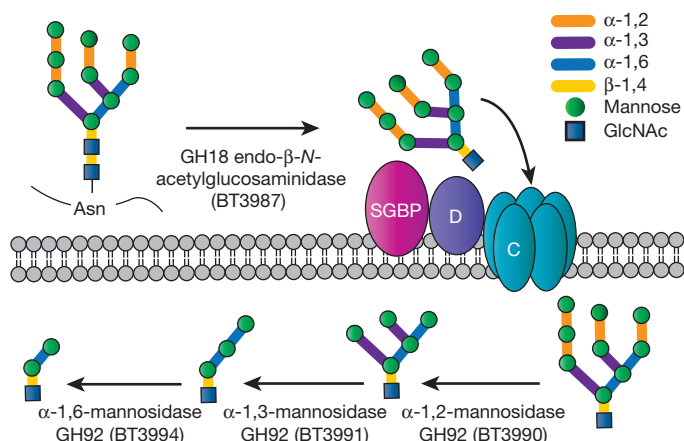


Figure 4 | Model of HMNG depolymerization by *B. thetaiotaomicron*. The interaction of mannose with the surface glycan binding protein (SGBP) is shown, while the binding of the SusD homologue (D) to the reducing end GlcNAc directs the N-glycan into the SusC-like (C) porin. In both the yeast mannan and the HMNG models the majority of glycan degradation occurs in the periplasm and not on the surface. The enzymes probably involved in the periplasmic degradation of the trisaccharide generated by the HMNG-PUL-encoded system are addressed in Supplementary Information section 4.3.

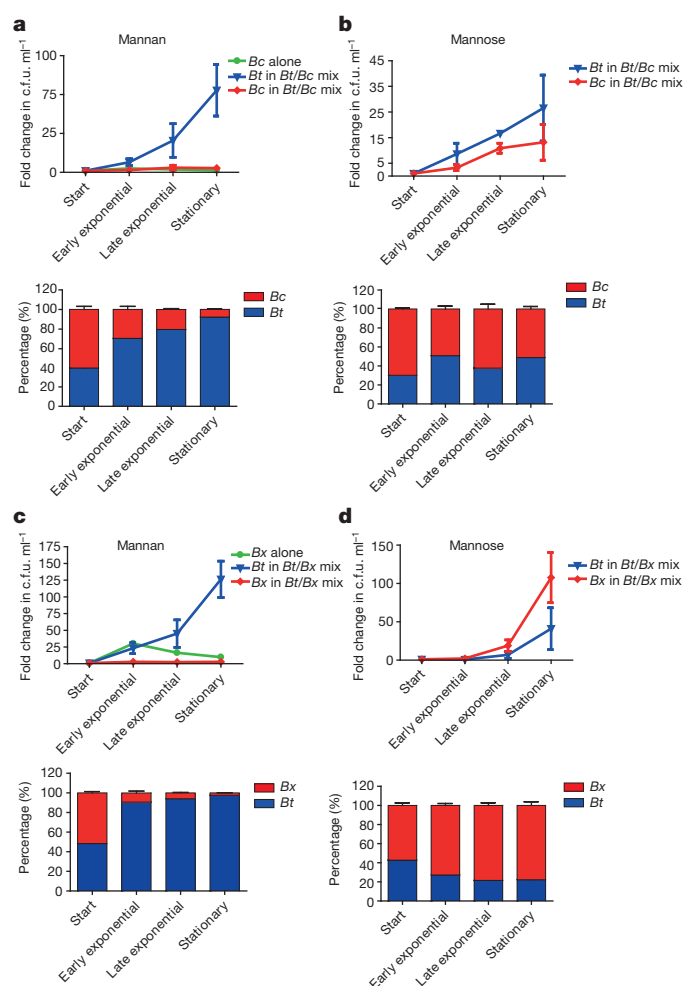


Figure 5 | *Bacteroides* co-culture sharing experiments. a–d, *B. thetaiotaomicron* was co-cultured with *Bacteroides cellulosilyticus* WH2 (a, b) or with *Bacteroides xylanisolvens* NLAE-zI-p352 (c, d) with either mannan (a, c) or mannose (b, d) as the sole carbon source. Each non-mannan user was also cultured on mannan independently. The upper graph in each panel depicts the c.f.u. ml⁻¹ of each strain, relative to the c.f.u. ml⁻¹ at inoculation. Total c.f.u. ml⁻¹ was determined by colony counts, and the proportion of each bacteria was determined by qPCR of marker genes from genomic DNA (shown in the lower graph of each panel). Error bars represent s.d. of three biological replicates. Bc, *B. cellulosilyticus*; Bt, *B. thetaiotaomicron*; Bx, *B. xylanisolvens*.

within widespread members of the human microbiota, thereby revealing the impact made by the historical domestication of yeast and other dietary fungi on the structure of this microbial consortium.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 24 January; accepted 22 October 2014.

- Arumugam, M. *et al.* Enterotypes of the human gut microbiome. *Nature* **473**, 174–180 (2011).
- Bäckhed, F., Ley, R. E., Sonnenburg, J. L., Peterson, D. A. & Gordon, J. I. Host-bacterial mutualism in the human intestine. *Science* **307**, 1915–1920 (2005).
- Arpaia, N. *et al.* Metabolites produced by commensal bacteria promote peripheral regulatory T-cell generation. *Nature* **504**, 451–455 (2013).
- Flint, H. J., Bayer, E. A., Rincon, M. T., Lamed, R. & White, B. A. Polysaccharide utilization by gut bacteria: potential for new insights from genomic analysis. *Nature Rev. Microbiol.* **6**, 121–131 (2008).
- Kau, A. L., Ahern, P. P., Griffin, N. W., Goodman, A. L. & Gordon, J. I. Human nutrition, the gut microbiome and the immune system. *Nature* **474**, 327–336 (2011).
- Round, J. L. & Mazmanian, S. K. The gut microbiota shapes intestinal immune responses during health and disease. *Nature Rev. Immunol.* **9**, 313–323 (2009).

- Martens, E. C., Koropatkin, N. M., Smith, T. J. & Gordon, J. I. Complex glycan catabolism by the human gut microbiota: the Bacteroidetes Sus-like paradigm. *J. Biol. Chem.* **284**, 24673–24677 (2009).
- El Kaoutari, A., Armougom, F., Gordon, J. I., Raoult, D. & Henrissat, B. The abundance and variety of carbohydrate-active enzymes in the human gut microbiota. *Nature Rev. Microbiol.* **11**, 497–504 (2013).
- Lombard, V., Golaconda Ramulu, H., Drula, E., Coutinho, P. M. & Henrissat, B. The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acids Res.* **42**, D490–D495 (2014).
- Konrad, A. *et al.* Immune sensitization to yeast antigens in ASCA-positive patients with Crohn's disease. *Inflamm. Bowel Dis.* **10**, 97–105 (2004).
- Mpofu, C. M. *et al.* Microbial mannan inhibits bacterial killing by macrophages: a possible pathogenic mechanism for Crohn's disease. *Gastroenterology* **133**, 1487–1498 (2007).
- Xu, J. *et al.* A genomic view of the human–*Bacteroides thetaiotaomicron* symbiosis. *Science* **299**, 2074–2076 (2003).
- Martens, E. C., Chiang, H. C. & Gordon, J. I. Mucosal glycan foraging enhances fitness and transmission of a saccharolytic human gut bacterial symbiont. *Cell Host Microbe* **4**, 447–457 (2008).
- Ballou, C. E., Ballou, L. & Ball, G. *Schizosaccharomyces pombe* glycosylation mutant with altered cell surface properties. *Proc. Natl Acad. Sci. USA* **91**, 9327–9331 (1994).
- Raschke, W. C., Kern, K. A., Antalis, C. & Ballou, C. E. Genetic control of yeast mannan structure. Isolation and characterization of mannan mutants. *J. Biol. Chem.* **248**, 4660–4666 (1973).
- Gregg, K. J. *et al.* Analysis of a new family of widely distributed metal-independent alpha-mannosidases provides unique insight into the processing of N-linked glycans. *J. Biol. Chem.* **286**, 15586–15596 (2011).
- Zhu, Y. *et al.* Mechanistic insights into a Ca²⁺-dependent family of α -mannosidases in a human gut symbiont. *Nature Chem. Biol.* **6**, 125–132 (2010).
- Rakoff-Nahoum, S., Coyne, M. J. & Comstock, L. E. An ecological network of polysaccharide utilization among human intestinal symbionts. *Curr. Biol.* **24**, 40–49 (2014).
- Ze, X., Duncan, S. H., Louis, P. & Flint, H. J. *Ruminococcus bromii* is a keystone species for the degradation of resistant starch in the human colon. *ISME J.* **6**, 1535–1543 (2012).
- Hehemann, J. H. *et al.* Transfer of carbohydrate-active enzymes from marine bacteria to Japanese gut microbiota. *Nature* **464**, 908–912 (2010).
- Larsbrink, J. *et al.* A discrete genetic locus confers xyloglucan metabolism in select human gut Bacteroidetes. *Nature* **506**, 498–502 (2014).
- Martens, E. C., Kelly, A. G., Taubin, A. S. & Brumer, H. The devil lies in the details: how variations in polysaccharide fine-structure impact the physiology and evolution of gut microbes. *J. Mol. Biol.* **426**, 3851–3865 (2014).
- Martens, E. C. *et al.* Recognition and degradation of plant cell wall polysaccharides by two human gut symbionts. *PLoS Biol.* **9**, e1001221 (2011).
- Everard, A., Matamoros, S., Geurts, L., Delzenne, N. M. & Cani, P. D. *Saccharomyces boulardii* administration changes gut microbiota and reduces hepatic steatosis, low-grade inflammation, and fat mass in obese and type 2 diabetic db/db mice. *MBio* **5**, e01011–e01014 (2014).

Supplementary Information is available in the online version of the paper.

Acknowledgements This work was supported by grants from the European Research Council (G.J.D., Glycopoise; H.J.G., No. 322820), The Wellcome Trust (H.J.G., WT097907AIA), BBSRC (M.J.T., G.J.D.; BB/G016127/1), US Department of Energy (DOE) Bioenergy Research Center (BESC) supported by the Office of Biological and Environmental Research in the DOE Office of Science (M.J.P.) and the National Institutes of Health (E.C.M. and T.J.T., GM090080). Gnotobiotic mouse experiments were supported by a subsidy from the University of Michigan Medical School Host Microbiome Initiative, Agriculture and Agri-Food Canada, AgriFlex (D.W.A., #2510), Canadian Institute of Health Research operating grant (A.B.B., MOP-68913), Australian Research Council; Mizutani Foundation (S.J.W.). We thank the staff of the Diamond Light Source for the provision of beamline facilities. We would also like to thank various members of iCaMB for providing the yeast strains used in this work. We were greatly saddened by the passing of C.Z. during the course of this work.

Author Contributions Enzyme characterization: F.C., M.J.T., J.L.M.-M. and D.W.A. Capillary electrophoresis: D.B., K.P. and W.V. E.C.L. created gene deletion strains and determined phenotypes with F.C. E.C.L., F.C. and A.R. performed enzyme localisation. F.C. and E.C.L. carried out the co-culturing experiments. Gene expression analysis: E.A.C., N.A.P. and E.C.M. Growth analysis on purified mannans and HMNG: E.C.L., N.A.P., K.U. and E.C.M. Characterization of HMNG binding proteins: Y.Z. Characterization of the *Abt3774* mutant: A.D. Phylogenetic reconstruction and metagenomic analysis: E.C.M. Gnotobiotic mouse experiments: E.A.C., N.A.P., N.T.P. and E.C.M. Purification of HMNG: T.J.T., B.S.H. and R.C. Isolation and genomic analysis of pig gut strains: T.A., C.J.Z. A.C. and G.S. performed NMR experiments on GH76 and M.J.P. on GT products. Z.H. and G.S. synthesized substrates. Crystallographic studies by A.J.T., G.J.D., M.D.S., A.B.B. and R.M. Experiments designed by F.C., E.C.L., G.J.D., S.J.W., D.W.A., E.C.M. and H.J.G. The manuscript was written primarily by H.J.G. and E.C.M. with contributions from S.J.W., G.J.D. and D.W.A. E.C.L. and E.C.M. prepared the figures.

Author Information The protein crystal structures reported in this study have been deposited under the following PDB accession codes: 4C1R (BT3783-Mg binary complex); 4C1S (BT3792); and 4UTF (BxGH99/Man-IFG/mannobiose ternary complex). Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to H.J.G. (harry.gilbert@ncl.ac.uk), E.C.M. (emartens@umich.edu) or D.W.A. (wade.abbott@agr.gc.ca).

METHODS

Producing recombinant proteins for biochemical assays. DNAs encoding the mature forms of the enzymes used in this study were amplified by PCR using appropriate primers. The amplified DNAs were cloned into NcoI/XhoI, NcoI/BamHI, NdeI/XhoI or NdeI/BamHI restricted pET21a or pET28a, as appropriate. The encoded recombinants generally contained a C-terminal His₆-tag, although, where appropriate, the His-tag was located at the N-terminus of the protein. The GH92 enzymes encoded by HMNG-PUL were cloned in a previous study¹⁷. To express the recombinant genes encoding the mannan degrading enzymes, *Escherichia coli* strains BL21(DE3) or TUNER, containing appropriate recombinant plasmids, were cultured to mid-exponential phase in Luria Bertani broth at 37 °C. This was followed by the addition of 1 mM (strain BL21(DE3)) or 0.2 mM (TUNER) isopropyl β-D-galactopyranoside (IPTG) to induce recombinant gene expression, and the culture was incubated for a further 5 h at 37 °C or 16 h at 16 °C, respectively. The recombinant proteins were purified to >90% electrophoretic purity by immobilized metal ion affinity chromatography using Talon, a cobalt-based matrix, and eluted with 100 mM imidazole, as described previously²⁵.

Producing recombinant proteins BT3783, BT3792 and BxGH99 for crystallization. pET28a expression vectors encoding mature BT3783 (residues 27 to 314) and BT3792 (residues 155 to 514) were cloned into pET28a via NheI and XhoI sites. The gene encoding the *B. xylanisolvens* homologue of GH99 (BxGH99) was constructed from synthesized oligonucleotide fragments (Genscript) and also subcloned into pET28a using NdeI and XhoI restriction sites. Plasmids encoding BT3783 and BT3792 were transformed into *E. coli* BL21(DE3) Star chemically competent cells, and grown in LB broth at 37 °C supplemented with 50 µg ml⁻¹ kanamycin. Production of recombinant BT3783 was induced by the addition of 0.2 mM IPTG at a culture OD₆₀₀ = 0.6, and incubation at 16 °C for 16 h. BT3792 was produced using the auto-induction method²⁶ by shaking inoculated 0.5-l cultures supplemented with 50 µg ml⁻¹ kanamycin for 36 h at 37 °C and then 48 h at 20 °C. *E. coli* BL21 (DE3) cells harbouring the BxGH99-encoding plasmid were cultured in 0.5 l ZYM-5052 auto-induction media²⁶, supplemented with 50 µg ml⁻¹ kanamycin, at 37 °C for 8 h, with induction occurring overnight at 16 °C.

To purify BT3783 and BT3792, cells were collected by centrifugation and ruptured by chemical lysis procedure at 4 °C. In brief, cells were resuspended in 25 ml of a solution consisting of 7% (w/v) sucrose, 50 mM Tris-HCl (pH 7.5), for 5 min. Lysozyme (Sigma-Aldrich; 10 mg) was then added and stirred for 10 min. A solution (50 ml) consisting of 0.6% (w/v) Triton-X, 0.6% (w/v) deoxycholate, 20 mM Tris-HCl (pH 7.5) was added and stirring continued, after which 5 mM MgCl₂ was added followed by addition of DNase (Sigma-Aldrich) to a final concentration of 8.5 µg ml⁻¹. The resulting solutions were then centrifuged for 45 min at 13,000g. Similarly, cells harbouring BxGH99 were collected and resuspended in 50 mM NaH₂PO₄ (pH 8.0), 300 mM NaCl, and lysed by sonication.

BT3783 clarified cell lysates were purified by nickel Sepharose affinity chromatography by stepwise elution with imidazole. Positive fractions identified by SDS gel electrophoresis were dialysed into 20 mM Tris (pH 8.0) and concentrated to 11.25 mg ml⁻¹ with an Amicon stirred ultrafiltration unit model 8200. BT3792 clarified cell lysates were purified using nickel-affinity chromatography, and positive fractions were buffer exchanged by dialysis into 25 mM Tris-HCl (pH 8.0) before further purification by anion exchange chromatography. Anion exchange purified BT3792 buffered in 25 mM Tris-HCl (pH 8.0), 500 mM NaCl, and 2 mM dithiothreitol was concentrated using a stirred cell Amicon with a 10 kDa cutoff to 20 mg ml⁻¹. For BxGH99 soluble lysate (isolated by centrifuging at 13,000g) was applied to a NiSO₄-charged 5 ml HiTrap chelating column (GE Healthcare), pre-equilibrated in the same buffer. Protein was eluted in an imidazole gradient, dialysed, concentrated, and further purified on an S75 16/60 gel filtration column (GE) pre-equilibrated in 25 mM Na-HEPES (pH 7.0), 100 mM NaCl, 1 mM dithiothreitol.

Binding studies. Affinity gel electrophoresis was used to screen for potential glycan binding proteins, following the method described in ref. 27 with the target polysaccharide at 1 mg ml⁻¹. The binding of proteins to their ligands was quantified by isothermal titration calorimetry (ITC), as described previously²⁷. Titrations were carried out in 50 mM Na-HEPES buffer, pH 7.5 at 25 °C. The reaction cell contained protein at 50 µM, while the syringe contained either the oligosaccharide at 10 mM or polysaccharide at 3–5 mg ml⁻¹. The titrations were analysed using Microcal Origin version 7.0 software to derive *n*, *K*_{av} and Δ*H* values, while Δ*S* was calculated using the standard thermodynamic equation, $RT \ln K_a = \Delta G = \Delta H - T\Delta S$.

Glycoside hydrolase and phosphatase assays. α-Mannosidase activity was determined by the continuous monitoring of mannose release using a linked enzyme assay system purchased from Megazyme International (mannose/fructose/glucose detection kit). The reaction was carried out at 37 °C in 50 mM Na-HEPES buffer (pH 7.0) containing 2 mM MgCl₂, 1 mM ATP and 1 mM NADP⁺, excess concentrations of linker enzymes (hexokinase, phosphoglucose isomerase and glucose-6-phosphate dehydrogenase) and 1 mg ml⁻¹ BSA. Through linker enzymes

glucose-6-phosphate, generated from the released mannose, is oxidized by glucose-6-phosphate dehydrogenase with concomitant reduction of NADP⁺ to NADPH, which was monitored at 340 nm using an extinction coefficient of 6223 M⁻¹ cm⁻¹. The polysaccharide substrates used were *S. cerevisiae* α-mannans from wild-type or appropriate mutant strains that produce variants of the polysaccharide (Extended Data Fig. 5); these mannans were purified from stationary-phase cultures of the yeast grown in yeast extract peptone dextrose medium, as described previously¹⁵. The α-1,6-mannooligosaccharides, which were also used as substrates, were generated as follows: 1 g α-mannan from the *S. cerevisiae* mutant mnn2 (comprises the α-1,6-mannan backbone with no side chains) was digested to completion with the endo-mannanase BT3792. The products were freeze-dried and the small manno-oligosaccharides (d.p. 2 to 5) were purified on two P2 Bio-gel columns set up in series, while the large oligosaccharides, with a d.p. of 6 to 8, were fractionated on two P4 Bio-gel columns also in series. The columns were run at 0.2 ml min⁻¹ in distilled water. The 5-ml fractions were evaluated by thin-layer chromatography (TLC) and those containing the same oligosaccharide were pooled. The activity of the endo-mannanases against α-mannans was determined in 50 mM Na-HEPES buffer (pH 7.5) containing an appropriate concentration of the polysaccharide (ranging from 0.1–6 mg ml⁻¹) and 1 mg ml⁻¹ BSA. Reactions were incubated for 30 min at 37 °C and, at regular time intervals, 500-µl aliquots were removed and the amount of reducing sugar was quantified deploying the dinitrosalicylic acid reagent²⁸ and a standard curve of mannose in the reaction conditions used. TLC was also used to provide a qualitative profile of the manno-oligosaccharides generated by the GH76 endo-mannanases from these reactions. Around 4 µl of the reaction was spotted on silica gel TLC plates and the plates were developed in butanol:acetic acid:water 2:1:1 and carbohydrate products detected by spraying with 0.5% orcinol in 10% sulfuric acid and heating to 100 °C for 10 min. Substrate-depletion assays were used to determine the activity of the endo-mannanases against α-1,6-mannooligosaccharides. In brief, 50 µM of the oligosaccharides in 50 mM sodium phosphate buffer (pH 7.5), containing 0.1 mg ml⁻¹ BSA (NaP buffer), was incubated with an appropriate concentration of enzyme (10 to 500 nM). Aliquots were removed at regular intervals for up to 1 h and, after boiling for 10 min to inactivate the enzyme, the amount of the substrate manno-oligosaccharide remaining was quantified by high-performance anion-exchange chromatography (HPAEC) using standard methodology. In brief, the reaction products were bound to a Dionex CarboPac PA-200 column equilibrated with 100 mM NaOH. Mannose and manno-oligosaccharides were eluted with a 0–200 mM sodium acetate gradient in 100 mM NaOH at a flow rate of 0.25 ml min⁻¹, using pulsed amperometric detection. The data were used to determine catalytic efficiency (*k*_{cat}/*K*_M) as described previously²⁹. To determine the activity of the GH99 endo-α-1,2-mannosidase BT3862, the enzyme was incubated with Man-α-1,3-Man-α-1,2-Man-α-1,2-Man-1-CH₃, and, in combination with the α-1,2-mannosidase BT3990, mannose release was monitored using the continuous assay described above (BT3990 is active on the product of BT3862, Man-α-1,2-Man-α-1,2-Man-1-CH₃). The activity of BT3862 against aryl-mannosidases was as described previously³⁰. A strategy deploying 2-aminobenzamide (2-AB)-labelled glycans at the reducing end, using the Sigma GlycoProfile 2-AB labelling kit as per the manufacturer's instructions, was used to determine the catalytic efficiency of BT3994. The enzyme (100 nM) was incubated with 5 µM Man₅GlcNAc₂ (generated by treating Man₅GlcNAc₂ with BT3990 an α-1,2-mannosidases) or Man₃GlcNAc₂ (generated from Man₅GlcNAc₂ by treating with BT3991, α-1,3-mannosidase¹⁷) in NaP buffer for up to 1 h. Aliquots were removed at regular intervals and the amount of glycan remaining was determined using HPAEC and a fluorometric detector. A similar strategy, employing labelled glycans in combination with HPAEC and fluorescence detection was used to measure the endo-β-N-acetylglucosaminidase activity of BT3987 against Man₅GlcNAc₂. The activity of the enzyme was also evaluated against glycoproteins using matrix-assisted laser desorption/ionization time of flight (MALDI-TOF) mass spectrometry as described below. BT2630 and BT3783 were assayed for phosphatase activity using the EnzChek Phosphate Assay kit to detect the release of phosphate from appropriate phosphorylated sugars. Prior to these assays the two proteins were treated with EDTA and the chelating agent was removed using a PD10 column. The assays were carried out in 50 mM Na-HEPES buffer (pH 7.5) containing 1 mM of MgCl₂ or another divalent ion. In this discontinuous assay aliquots were removed at regular intervals, the enzyme was inactivated by boiling and phosphate release determined.

The specificity of enzymes against phosphorylated and neutral high mannose N-glycans was assessed by capillary electrophoresis. The methodology was essentially that described in ref. 31. The two substrates were MNN4 (phosphorylated high mannose N-glycans released from glycoproteins expressed by a genetically modified strain of *Yarrowia lipolytica*) and N-glycan from RNAase B (primarily high mannose N-glycans), which were released from their respective glycoproteins with PNGase F. The N-glycans were labelled at the reducing end with 8-aminopyrene-1,3,6-trisulfonic acid³², and incubated overnight at 30 °C with

~2 μM of enzyme in 10 mM Na-Hepes buffer (pH 7.0) containing 2 mM CaCl_2 . The reactions were analysed by capillary electrophoresis with laser-induced fluorescence detection (CE-LIF) using an ABI 3130 capillary DNA sequencer as described previously³².

Mannosyl-transferase assays. BT3775 and BT3776 at 30 μM were screened for activity using mannose and all possible α -mannobiose acceptors at a concentration of 10 mM. Reactions were performed in 10 mM Na-Hepes buffer (pH 7.5), 10 mM MnCl_2 , 50 mM GDP-Mannose at 37 °C over time course intervals of 0 min, 5 min, 15 min, 30 min, 60 min and 16 h. Reactions were stopped by boiling for 10 min. To explore the synergy between the two GT32 glycosyltransferases, reactions were performed with BT3775 and BT3776 simultaneously using mannose as an acceptor, and in a step-wise progression with BT3775 and mannose as the acceptor, followed by BT3776 on the purified products of the initial BT3775 reaction. Enzyme concentrations were maintained in both conditions at 30 μM .

MALDI-TOF mass spectrometry of reaction products. To permethylate *N*-glycan fragments and manno oligosaccharides appropriate enzyme reactions were freeze-dried and suspended in dry dimethylsulfoxide (DMSO; 200 μl). Oligoglycans were per-O-methylated using standard methods³³. Freshly prepared sodium hydroxide base in dry DMSO (300 μl) and then iodomethane (150 μl) was added to each sample. The tube was purged with nitrogen and vortexed vigorously. Permethylated oligosaccharides (PMOs) were extracted in 2 ml water. After removing excess iodomethane (CH_3I), 2 ml dichloromethane was added and vortexed to extract lipophilic components. The aqueous layer was removed after centrifugation. This was repeated five more times. After the final rinse, the dichloromethane layer was transferred into a fresh tube and dried down under N_2 stream. The samples were then dehydrated and dissolved in 50% aqueous methanol and loaded onto C18 reverse phase resin (100 mg ml^{-1} , Resprep SPE Cartridges) and eluted with acetonitrile (AcN). AcN washes were dried down under N_2 and dissolved in 20 μl methanol. Equal volumes of the sample and 2% DHBA in aqueous methanol, which was used as the matrix, were combined in a separate microfuge tube and loaded on the MALDI-TOF mass spectrometry plate. MALDI-TOF mass spectrometry analyses were performed with an AB4700 MS instrument (Applied Biosystems) in positive ion, reflector mode with 100 shots taken per spectrum. PMO mass values were queried online (<http://www.expasy.org/tools/glycomod/>) by GlycoMod tool that can predict possible oligosaccharide structures that occur on proteins from their experimentally determined masses.

NMR analysis of glycosyltransferase reaction products. The manno oligosaccharide products were separated by size exclusion chromatography on a Superdex-75 HR10/30 column using a Dionex Ultimate 3000 HPLC equipped with a Shodex RI-101 refractive index detector. The column was eluted with water and fractions were collected and freeze-dried. The lyophilized fraction containing the trisaccharide was dissolved in D_2O (0.3 ml, 99.9%; Cambridge Isotope Laboratories) and one- and two-dimensional NMR spectra were recorded at 298 K with a Varian Inova NMR spectrometer operating at 600 MHz equipped with a 3 mm NMR cold probe. The homonuclear (gCOSY, TOCSY and NOESY) and the heteronuclear (gHSQC and gHMBC) experiments were recorded using standard Varian pulse programs. Chemical shifts were measured relative to internal acetone (δ_{H} 2.225). Data were processed using MestReNova software (Universidad de Santiago de Compostela, Spain). Interglycosidic scalar couplings observed in the gHMBC spectra were used to determine the sequence and glycosidic linkages of the mannosyl residues in the α -Man-(1,3)-[α -Man-(1,6)]-Man trisaccharide product.

Linkage analysis of GT32 glycosyltransferase manno oligosaccharide products. Overnight incubations were boiled for 5 min and then centrifuged to pellet denatured protein. A 200 μg aliquot of the sample was suspended in 200 μl DMSO and magnetically stirred for 5 days. The sample was then permethylated according to the method in ref. 34. Two hundred microlitres of the NaOH base were added and, after 10 min, 100 μl CH_3I were added and the sample was stirred for 40 min. An additional 200 μl base and 100 μl CH_3I were then added and stirring was continued for 40 min. The reaction was worked up by addition of 2 ml H_2O , removal of excess CH_3I by sparging with N_2 , and CH_2Cl_2 extraction. The permethylated material was hydrolyzed using 2 M trifluoroacetic acid (TFA) (2 h in a sealed tube at 121 °C), reduced with NaBD₄, and acetylated using Ac_2O /TFA. The resulting partially methylated alditol acetates were analysed on a Hewlett Packard 5890 GC interfaced to a 5970 MSD (mass selective detector, electron impact ionization mode); separation was performed on a 30 m Supelco 2330 bonded phase fused silica capillary column.

Stereochemistry of endo- α -1,6-mannanases. A solution of BT3792 (25 mg, approximately 400 μmol) in buffered D_2O (0.25 ml, 50 mM phosphate-citrate, 200 mM NaCl, pD 6.0 (deuterium ion concentration is 1 μM)) was added to a solution of 4-nitrophenyl α -D-mannopyranosyl-1,6- α -D-mannopyranoside (4.1 mg, 0.0086 mmol) in D_2O (0.75 ml, 50 mM phosphate-citrate, 200 mM NaCl, pD 6.0) at 22 °C. The course of the reaction was monitored by ^1H NMR (500 MHz) to identify the stereochemistry of the reaction by analysing the relative timing of the

appearance of the anomeric α - and β -proton signals. Signals were assigned by two-dimensional NMR analysis (HSQC).

Bacteroides culture and whole-cell assays. *B. thetaiotaomicron* was cultured anaerobically at 37 °C in minimal media containing an appropriate carbon source, or in TYG (tryptone yeast extract glucose medium) as described previously²¹. Growth curves presented in the paper are averages of six biological replicates. *B. thetaiotaomicron* was grown in 5 ml minimal media on 0.5% w/v *S. cerevisiae* mannan (Sigma) or glucose as the sole source to mid exponential phase ($\text{OD}_{600\text{nm}}$ 0.6–0.8). Cells were harvested by centrifugation, 5000g for 10 min at room temperature and washed in 5 ml PBS (pH 7.1) before being resuspended in 500 μl PBS. Cells (50 μl) were assayed against yeast mannan (10 mg ml^{-1}) at 37 °C for 16 h. Assays were analysed by thin layer chromatography, 5 μl of each sample was spotted onto silica plates and resolved in butanol/acetic acid/water buffer. The plates were dried and sugars visualized by orcinol/sulfuric acid heated to 70 °C.

Cellular localization. Cultures of *B. thetaiotaomicron* (100 ml) were grown in minimal media on yeast mannan (0.5% w/v) as a sole carbon source, to mid exponential growth phase ($\text{OD}_{600\text{nm}}$ 0.6–0.8). Cells were harvested by centrifugation and washed in 10 ml PBS before being resuspended in 5 ml of the buffer. The cells were split into four 1 ml aliquots. To 3 of the aliquots 2 mg ml^{-1} Proteinase K was added and incubated at 37 °C for 1–16 h, the fourth sample was left as an untreated control also for 16 h. Following incubation with the protease the samples were centrifuged at 5000g for 10 min and the supernatant discarded. The cell pellets were resuspended in 1 ml PBS and the proteins precipitated by the addition of 200 μl trichloroacetic acid and incubation on ice for 30 min. The precipitated proteins were pelleted by centrifugation and washed 4 times in 1 ml ice-cold acetone. The protein pellets were resuspended in 250 μl Laemmli buffer and subjected to SDS/PAGE. Gels were transferred to Whatman Protran BA 85 nitrocellulose membrane. Proteins of interest were detected using anti-sera raised against the corresponding protein. The secondary antibody used was a chicken anti-rat conjugated to horseradish peroxidase. Antibodies were detected by chemi-luminescence using Biorad Clarity Western ECL Substrate.

Immunofluorescence microscopy. *B. thetaiotaomicron* suspensions ($\text{OD}_{600\text{nm}}$ = 0.8) in PBS (pH 7.0) were applied to clean Eppendorf tubes, fixed by an equal volume of 2 \times formalin (9% formaldehyde in PBS), and rocked for 90 min at 25 °C. The cells were then pelleted by centrifugation for 3 min at 7000g and washed twice with 1 ml of PBS. The bacterial cell pellet was resuspended in 1 ml of blocking solution (2% goat serum, 0.02% NaN_3 in PBS) and incubated at 4 °C for 16 h. After incubation cells were centrifuged again at 7000g and the supernatant discarded. For labelling, the bacteria were incubated with 0.5 ml of primary rat IgG (1/500 dilution of IgG in blocking solution) for 2 h at 25 °C. The cells were then pelleted, by centrifugation, washed in 1 ml of PBS and resuspended in 0.4 ml goat anti-rat IgG Alexa-Fluor 488 (Sigma), diluted 1/500 in blocking solution, and incubated 1 h at 25 °C in the dark. The cells were then pelleted, washed with PBS and resuspended in 50 μl of PBS containing ProLong Gold antifade reagent (Life Technologies). Labelled bacterial cells were mounted onto glass slides and secured with coverslips. Fluorescence was visualized using a Leica SP2 UV microscope (Leica Microsystems, Heidelberg, GmbH) with $\times 63$ NA 1.32 lens. Alexa-Fluor 488-labelled bacteria were viewed under an ultraviolet microscope view and compared with bright-field phase-contrast of the same image.

Constructing mutants of the yeast mannan and HMNG PULs in *B. thetaiotaomicron*. The inactivation of the HMNG-PUL extra-cellular factor regulator was described previously¹³. The other mutants deployed in this study, PUL knock-outs and single gene clean deletions, were introduced by allelic exchange using the pExchange vector as described in ref. 35 The use of quantitative reverse transcriptase PCR to quantify appropriate transcripts followed the methods described in ref. 13.

Gnotobiotic mice experiments. All animal experiments were approved by the University Committee on Use and Care of Animals at the University of Michigan and were supervised by a veterinarian. Germ-free mouse experiments were conducted in a total of 24, 6–8-week-old male and female Swiss Webster mice (each was considered to be a single biological replicate in its respective experiment and treatment group). Mice were randomly assigned into groups by a technician who was not familiar with the project. The investigators were not blinded to the identities of the treatment groups during the experiment and no data were excluded from the final analysis. For *in vivo* gene expression studies, mice were randomly grouped into three groups containing three animals each, and then subjected to three different dietary regimens. Two groups were fed a gamma-irradiated custom diet (glycan-free diet) that contained only glucose as the available carbohydrate as well as cellulose, a polysaccharide that cannot be degraded by *B. thetaiotaomicron*, as a non-digestible fibre supplement³⁶. One of the two groups maintained on this diet was provided with purified α -mannan (1% w/v) in drinking water (Harlan-Teklad). The third group was fed a custom version of the glycan-free diet in which the glucose was replaced with dried/crumbled bread (50% w/w of final diet) that

had been produced using yeast as a leavening agent (Zingerman's Bake House, Ann Arbor, MI). Mice were pre-fed on each dietary condition for five days, colonized with *B. thetaiotaomicron* by oral gavage ($\sim 10^8$ c.f.u. per animal) and maintained for an additional 5 days before euthanizing and collection of caecal contents for gene expression analysis. For the *in vivo* competition experiment, wild-type and mannan PUL triple mutant strains were each labelled with a unique 24 base pair (bp) oligonucleotide tag, which is contained in a pNBU2-based chromosomal integration vector and quantifiable by qPCR, as previously described¹⁵. Each strain was grown overnight in TYG medium and combined at approximately equal amounts before being gavaged into mice as described above. For competition experiments, five mice were used in each group (chosen based on the sample size used in previous studies to observe significant changes in competitive index^{18,29}) and fed varying regimens of the three diets described above for a total period of 38 days. All germfree mice used in the competition experiment were pre-fed the glycan-free diet for 1 week before colonization.

Bacteroides co-culturing experiments. *B. thetaiotaomicron*, *B. xylanisolvens* NLAE-zl p352 and *B. cellulosilyticus* WH2 were cultured in minimal media containing glucose (0.5% w/v) as a carbon source to mid exponential growth phase (OD₆₀₀ 0.7–0.8). Cells were collected by centrifugation and washed twice with 5 ml of PBS to remove any residual glucose. The washed cells were resuspended in 5 ml of fresh minimal media with no carbon source. Minimal media (10 ml) containing mannose or yeast mannan as the carbon source was inoculated with equal volumes of *B. thetaiotaomicron*/*B. xylanisolvens* or *B. thetaiotaomicron*/*B. cellulosilyticus*. A control culture of minimal media containing yeast mannan was also inoculated with either *B. xylanisolvens* or *B. cellulosilyticus*. All cultures were set up in triplicate. Samples (1 ml) were taken at the point of inoculation, early exponential, late exponential and stationary phase of growth. Serial dilutions of each sample were plated on to rich media and incubated for 2 days before colony counts were recorded. Proportions of *B. thetaiotaomicron*, *B. xylanisolvens* and *B. cellulosilyticus* per sample were determined by quantitative PCR from genomic DNA using unique marker genes for each strain.

Searches of human gut metagenomic data sets for mannan PULs. A search of 250 different healthy and diseased (Crohn's disease and ulcerative colitis) human metagenomic samples was conducted exactly as described previously³⁶. Probes corresponding to the three *B. thetaiotaomicron* mannan PULs were first checked against database of complete and draft microbial genome sequences to identify regions that were present in species other than *B. thetaiotaomicron*, or the porcine *B. xylanisolvens* strains in the case of MAN-PUL1. In the case of MAN-PUL2 and MAN-PUL3, the entire sequence was determined to be specific for sequenced *B. thetaiotaomicron* isolates; whereas, MAN-PUL1 was trimmed to a region containing 15,059 kilobases (kb) that were unique to *B. thetaiotaomicron* and porcine *B. xylanisolvens*. The three probes used correspond to the following base pair coordinates in the *B. thetaiotaomicron* VPI-5482 genome sequences: MAN-PUL1 (3262908..3277966), MAN-PUL2 (4893415..4928241) and MAN-PUL3 (5012915..5031551).

Crystallization. BT3783 and BT3792 were crystallized using the sitting drop vapour diffusion method at 18 °C. BT3783 crystals were obtained by mixing equal volumes of purified, recombinant BT3783 protein at a concentration of 11.25 mg ml⁻¹ with mother liquor solution consisting of 15% (w/v) polyethylene glycol 4,000, 5% (+/-) 2-methyl-2,4-pentanediol (v/v), 3 mM D-mannose, and 200 mM MgCl₂. Crystals of BT3792 were obtained by mixing equal volumes of purified recombinant BT3792 at a concentration of 20 mg ml⁻¹ buffered in 25 mM Tris-HCl (pH 8.0), 500 mM NaCl, 2 mM dithiothreitol with a reservoir solution consisting of 5% glycerol (v/v), 24% (w/v) polyethylene glycol monomethyl ether (average *M_w*, 2,000), 0.25 M sodium acetate, and Bis-Tris-HCl (pH 5.5). Large, thin plate crystals of BT3792 developed over a period of 5 days to a week. BT3792 crystals were cryoprotected in crystallization solutions supplemented with 25% ethylene glycol and cryo-cooled directly in a N₂ stream at -160 °C before diffraction experiments.

BxGH99 crystals were grown at 19 °C using hanging-drop vapour diffusion with equal volumes of protein (30 mg ml⁻¹) and reservoir solution consisting of 0.1 M sodium acetate pH 5.1, 20% w/v polyethylene glycol monomethyl ether (average *M_w*, 2,000), 2.0% low molecular weight poly-γ-glutamic acid. Ligand complex formation as a ternary with Man-isofagamine and α-1,2-mannobiose, was achieved by soaking native BxGH99 crystals in mother liquor supplemented with approximately 10 mM of respective ligand solutions for a period of 30 min before flash cooling in liquid N₂. Crystals were cryo-protected by the stepwise addition of ethylene glycol, supplemented with appropriate ligand(s), to a final concentration of 20% v/v.

Data collection, structure solution and refinement. Diffraction data from both BT3783 and BT3792 crystals were collected at the Canadian Lightsource beamline 08B1-1 (CMCF-BM) in oscillation mode ($\lambda = 0.97961$ and 0.98005 Å, with data collected at 100 K, respectively). The BT3783 and BT3792 data sets were processed with iMosflm^{37,38} and XDS³⁹, respectively, and both were scaled with

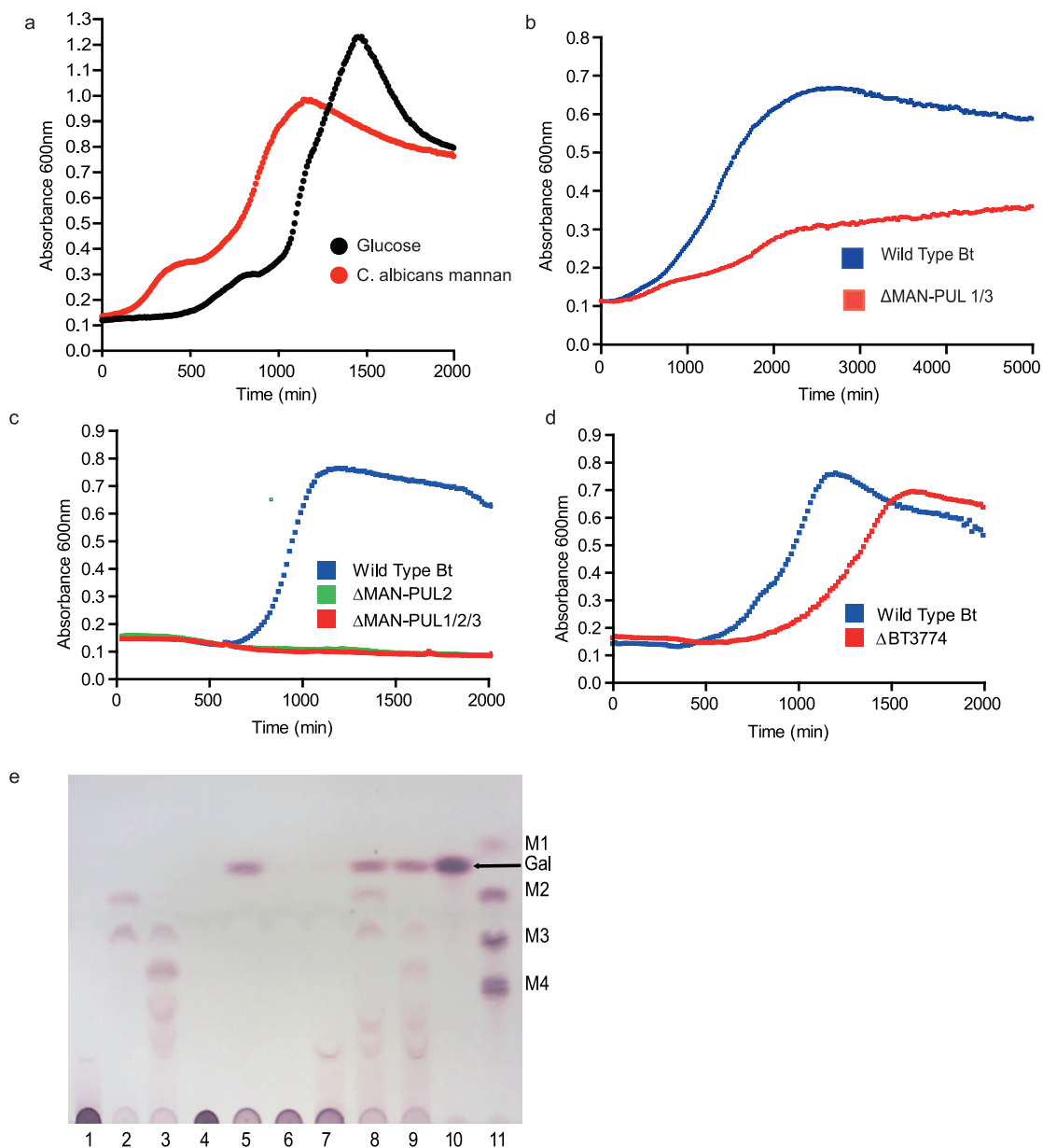
SCALA³⁸. BT3783 and BT3792 structures were solved by molecular replacement with the program Phaser MR⁴⁰ using the coordinates of apo-BT3783 (PDB: 3MPR) and the coordinates of Lin0763 from *Listeria innocua* (PDB: 3K7X). The asymmetric unit of BT3783 and BT3792 contain four and two protein molecules, respectively. The structures were subsequently improved with cycles of manual building with COOT⁴¹ and refinement with REFMAC³⁸. Five per cent of the reflections were flagged as 'free' to monitor the refinement process. The structure of BT3783 extends from residues 26–309 with a magnesium ion coordinated in the putative active centre. The structure of BT3792 runs continuously from residues 151–525. Due to a lack of continuous density corresponding to this polypeptide region, residues 274–290 are absent in the model of BT3783. BT3792 structural refinement required amplitude correction for twinning. The final refined structures were validated with the aid of Molprobity⁴². Final BT3783 and BT3792 structures featured, respectively, 94.8 and 99.6% of all modelled residues within the Ramachandran 'favoured' region, with a further 4.2 and 0.4% within the 'additional allowed' region.

Diffraction data for the BxGH99-Man-IFG-mannobiose ternary complex were collected at beamline I03 ($\lambda = 0.97625$ Å, data collected at 100 K) of the Diamond Light Source, Didcot, UK. The data sets were processed using XDS³⁹ and AIMLESS⁴³ (also within CCP4³⁸). Phases were derived from a previously solved (native) BxGH99 atomic model, with initial refinement conducted using REFMAC³⁸. Calculated electron density maps were visually assessed for evidence of ligand-binding, with subsequent completed atomic models, featuring appropriately assigned ligand coordinates, refined to convergence through numerous cycles of REFMAC³⁸ and additional manual correction using COOT⁴¹. The final BxGH99 ternary structure featured, respectively, 96.5 and 96.9% of all modelled residues within the Ramachandran 'favoured' region, with a further 2.3 and 1.9% within the 'additional allowed' region. Data collection and refinement statistics for all structures are presented in Supplementary Table 6.

Experimental group size and statistical analysis of data. For all quantitative enzyme assays at least three technical replicates were performed unless specifically stated in the Table legends. Standard errors of the mean are stated for kinetic parameters in Supplementary Tables 2–4. For Supplementary Table 5 the kinetic parameters are reported \pm the error of the fit of the data using linear regression. For isothermal titration calorimetry experiments three technical replicates were performed for each titration. Standard errors of the mean are shown in Supplementary Table 5. In the gnotobiotic mice experiments displayed in Fig. 2a five mice were assigned to each diet and the error bars for the two bacterial strains represent standard errors of the mean. Statistical significance of the persistence of different strains of *B. thetaiotaomicron* in the same mice was analysed by unpaired Student's *t*-test. For the microbial sharing experiment detailed in Fig. 4 the error bars represent standard deviations of the mean from three biological replicates.

25. Charnock, S. J. *et al.* Key residues in subsite F play a critical role in the activity of *Pseudomonas fluorescens* subspecies *cellulosa* xylanase A against xylooligosaccharides but not against highly polymeric substrates such as xylan. *J. Biol. Chem.* **272**, 2942–2951 (1997).
26. Studier, F. W. Protein production by auto-induction in high density shaking cultures. *Protein Expr. Purif.* **41**, 207–234 (2005).
27. Szabo, L. *et al.* Structure of a family 15 carbohydrate-binding module in complex with xylopentaose. Evidence that xylan binds in an approximate 3-fold helical conformation. *J. Biol. Chem.* **276**, 49061–49065 (2001).
28. Miller, G. L. Use of dinitrosalicylic acid reagent for determination of reducing sugar. *Anal. Chem.* **31**, 426–428 (1959).
29. Charnock, S. J. *et al.* The topology of the substrate binding clefts of glycosyl hydrolase family 10 xylanases are not conserved. *J. Biol. Chem.* **273**, 32187–32199 (1998).
30. Thompson, A. J. *et al.* Structural and mechanistic insight into N-glycan processing by endo-α-mannosidase. *Proc. Natl Acad. Sci. USA* **109**, 781–786 (2012).
31. Stewart, T. S., Mendershausen, P. B. & Ballou, C. E. Preparation of a mannopentaose, mannohexaose, and mannoheptaose from *Saccharomyces cerevisiae* mannan. *Biochemistry* **7**, 1843–1854 (1968).
32. Laroy, W., Contreras, R. & Callewaert, N. Glycome mapping on DNA sequencing equipment. *Nature Protocols* **1**, 397–405 (2006).
33. Ciucanu, I. Per-O-methylation reaction for structural analysis of carbohydrates by mass spectrometry. *Anal. Chim. Acta* **576**, 147–155 (2006).
34. Anumula, K. R. & Taylor, P. B. A comprehensive procedure for preparation of partially methylated alditol acetates from glycoprotein carbohydrates. *Anal. Biochem.* **203**, 101–108 (1992).
35. Koropatkin, N. M., Martens, E. C., Gordon, J. I. & Smith, T. J. Starch catabolism by a prominent human gut symbiont is directed by the recognition of amylose helices. *Structure* **16**, 1105–1115 (2008).
36. Larsbrink, J. *et al.* A discrete genetic locus confers xyloglucan metabolism in select human gut Bacteroidetes. *Nature* **506**, 498–502 (2014).
37. Leslie, A. W. & Powell, H. In *Evolving Methods for Macromolecular Crystallography* (eds Read, J. R. & Sussman, J. L.) Ch. 4, 41–51 (Springer, 2007).
38. Winn, M. D. *et al.* Overview of the CCP4 suite and current developments. *Acta Crystallogr. D* **67**, 235–242 (2011).
39. Kabsch, W. Xds. *Acta Crystallogr. D* **66**, 125–132 (2010).

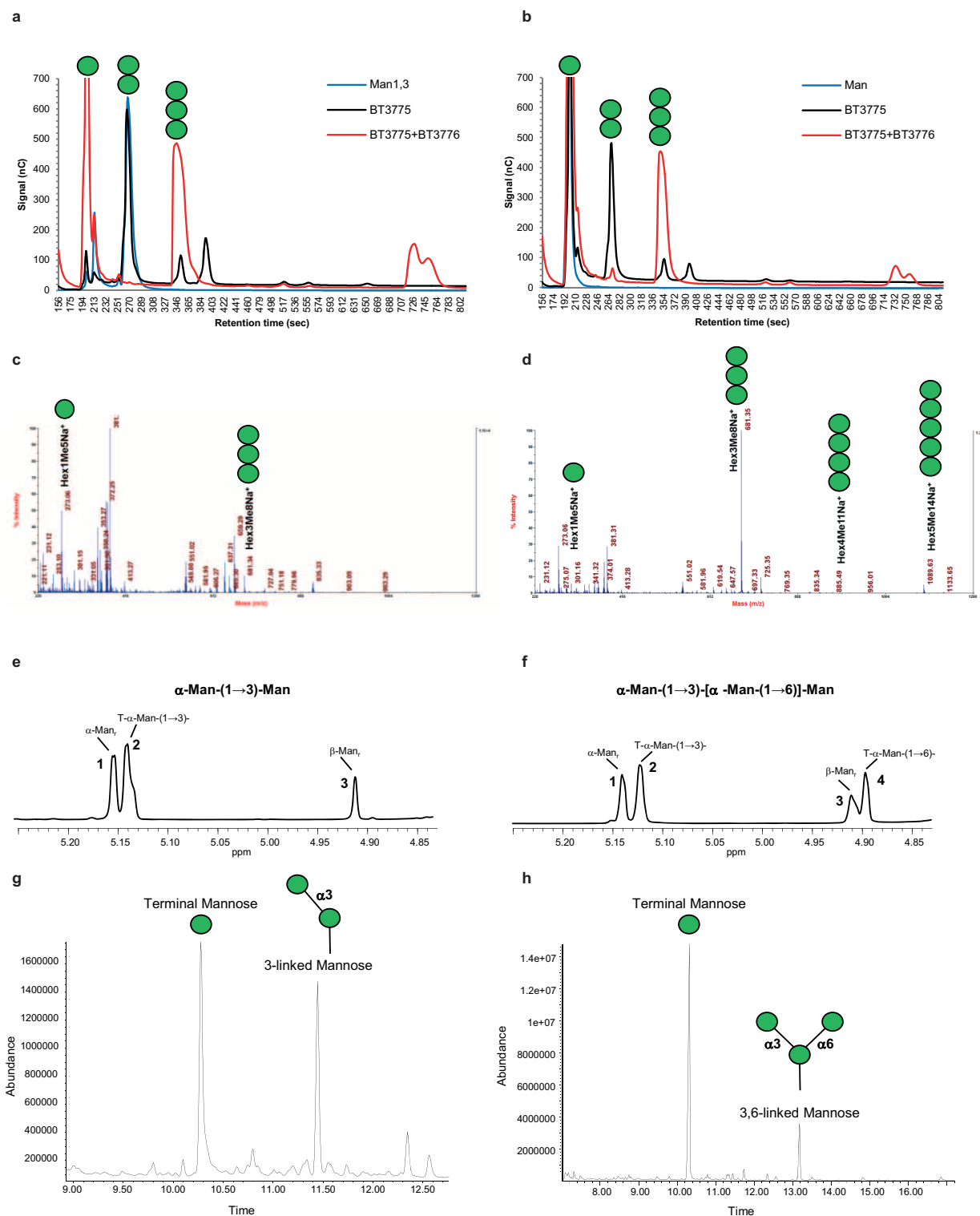
40. McCoy, A. J. *et al.* Phaser crystallographic software. *J. Appl. Crystallogr.* **40**, 658–674 (2007).
41. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. D* **66**, 486–501 (2010).
42. Chen, V. B. *et al.* MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr. D* **66**, 12–21 (2010).
43. Evans, P. R. & Murshudov, G. N. How good are my data and what is the resolution? *Acta Crystallogr. D* **69**, 1204–1214 (2013).



Extended Data Figure 1 | The role of specific *B. thetaiotaomicron* PULs and enzymes in utilization of mannan from *S. cerevisiae* and other yeast species.

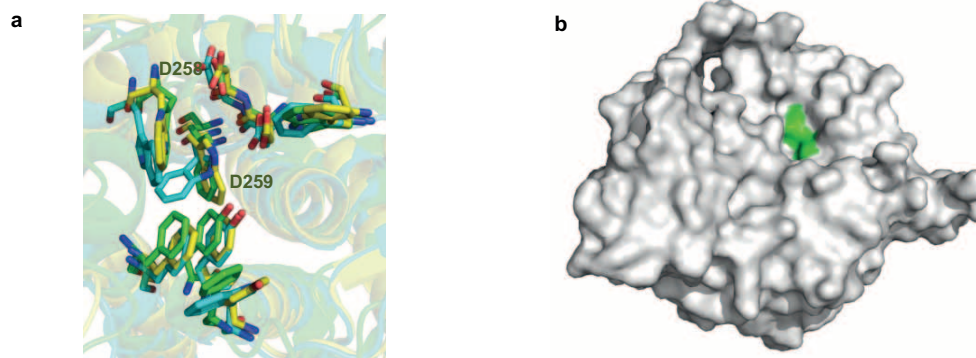
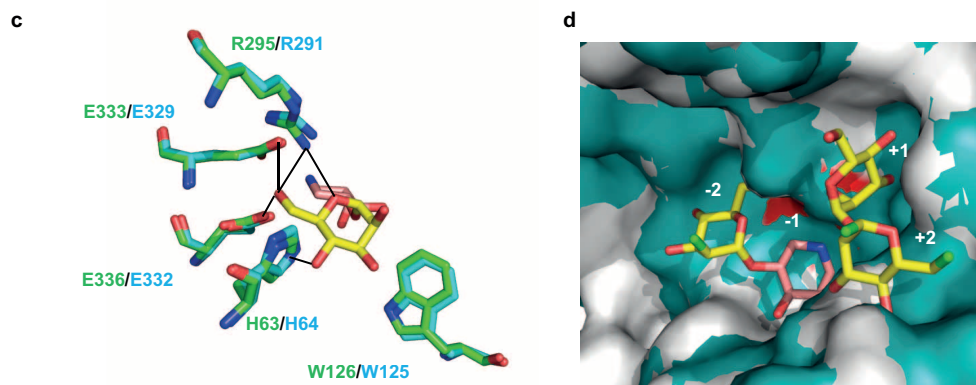
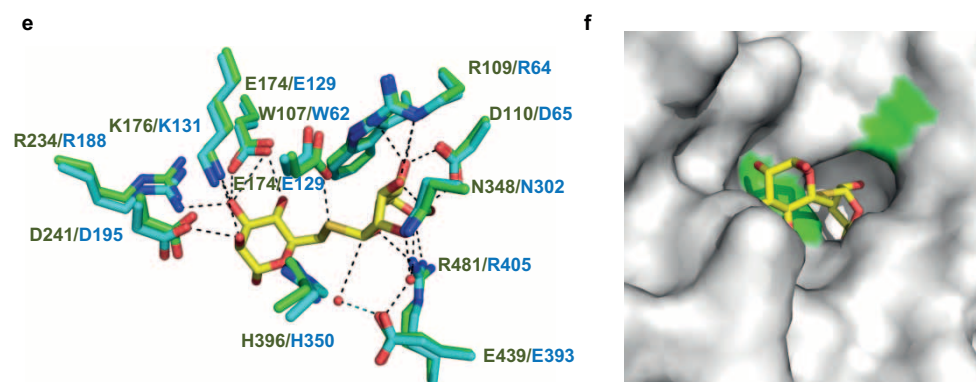
a, Growth of wild-type *B. thetaiotaomicron* on *Candida albicans* mannan and glucose. **b,** Growth of wild-type *B. thetaiotaomicron* and the mutant lacking MAN-PUL1 and MAN-PUL3 (Δ MAN-PUL1/3) on *Schizosaccharomyces pombe* α -mannan. **c,** Growth of wild-type *B. thetaiotaomicron*, and the *B. thetaiotaomicron* mutants lacking MAN-PUL2 (Δ MAN-PUL2), or all three mannan PULs (Δ MAN-PUL1/2/3) on *S. cerevisiae* α -mannan. **d,** The growth profile of wild-type *B. thetaiotaomicron* and the *B. thetaiotaomicron* mutant lacking *bt3774* (Δ bt3774) on *S. cerevisiae* mannan. In panels **a**, **b**, **c** and **d**, each point on the growth curve represents the mean of three biological replicates.

e, Enzymes at 1 μ M at 37 $^{\circ}$ C were incubated with either undecorated α -1,6-mannan (derived from *mn2* mutant of *S. cerevisiae*) (lanes 1–3) or mannan from *S. pombe* (lanes 4–9). Lanes 1 and 4, the mannans incubated in the absence of the enzymes; lanes 2 and 6, mannans incubated with the periplasmic mannanase BT3782; lanes 3 and 7, mannans incubated with the surface mannanase BT3792; lane 5, *S. pombe* mannan incubated with the GH97 α -galactosidase BT2620; lanes 8 and 9, *S. pombe* mannan incubated with BT2620/BT3782 and BT2620/BT3792, respectively. Lane 10 galactose standard; lane 11 α -1,6-mannooligosaccharides: mannose (M1), mannobiose (M2), mannotriose (M3) and mannotetraose (M4).

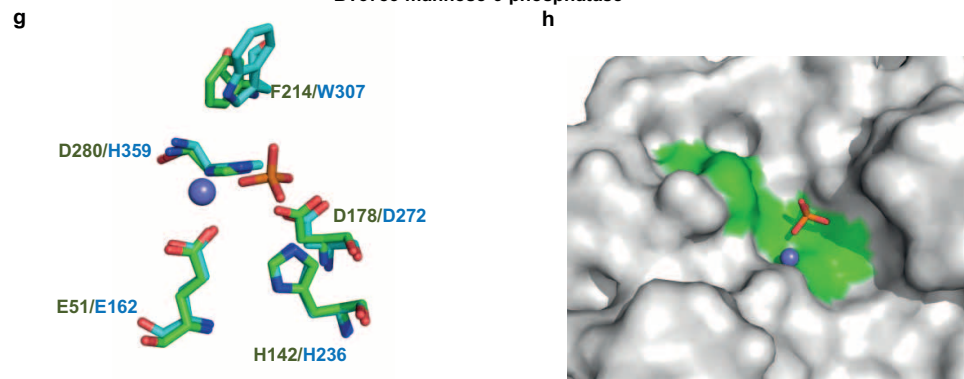


Extended Data Figure 2 | Product profiling of GT32 glycosyltransferases encoded by MAN-PUL2. **a**, HPAEC of biosynthetic reactions using mannose as an acceptor and GDP- α -Man as the donor. Mannobiose is formed in the presence of BT3775 (black) and mannotriose with BT3775 and BT3776 (red). The blue trace is a mannose standard. **b**, HPAEC of biosynthetic reactions with α -1,3-mannobiose as the acceptor and GDP- α -Man as the donor. BT3775 is not capable of extending mannobiose (black). In the presence of BT3775 and BT3776 mannotriose is produced (red). The blue trace is an α -1,3-mannobiose standard. **c**, **d**, MALDI-TOF analysis of the reaction products of BT3775 + BT3776 using mannose (c) and α -1,3-mannobiose (d) as an acceptor. **e**, NMR analysis of the α -1,3-mannobiose substrate. Peaks 1 and 3

correspond to the α -anomer and β -anomer of the mannose at the reducing end, respectively; peak 2 corresponds to the terminal α -mannosyl residue linked to O3 of the mannose at the reducing end. **f**, NMR analysis of the α -1,3, (α -1,6)-mannotriose BT3776 product. The numbering of the peaks are the same as in e. Peak 4 corresponds to the terminal α -mannosyl residue linked to O6 of the 3,6-linked mannose at the reducing end. **g**, Alditol-acetate linkage analysis of mannobiose produced by BT3775 from mannose. **h**, Alditol-acetate linkage analysis of branched (α -1,3), (α -1,6)-mannotriose produced by BT3776 from α -1,3-mannobiose. The green circles indicate the mannose residues present in carbohydrates identified by HPAEC, MALDI-TOF and NMR.

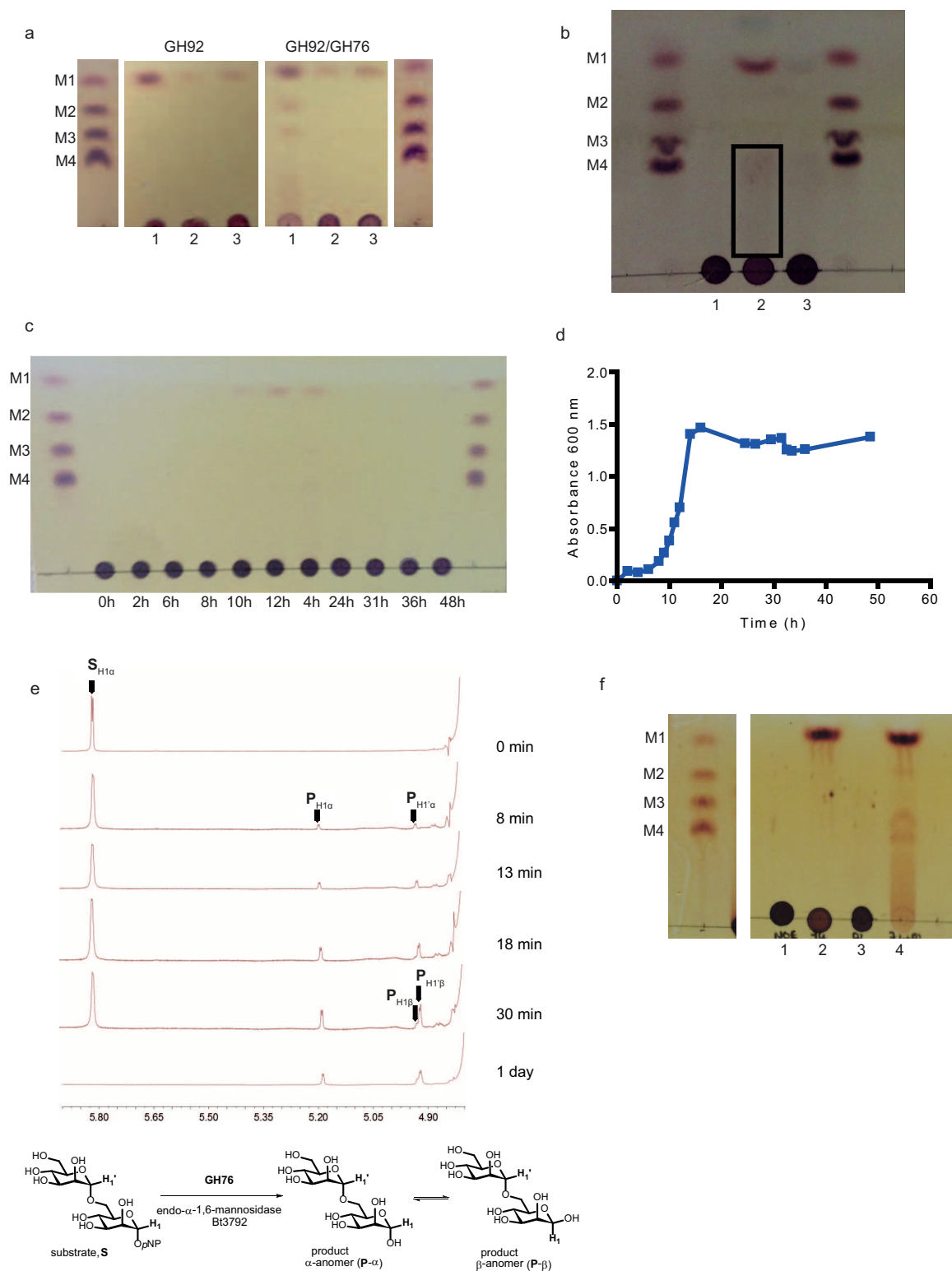
BT3792 GH76 endo- α 1,6-mannanaseBT3862 GH99 endo- α 1,2-mannosidaseBT3781 GH125 exo- α 1,6-mannosidase

BT3783 mannose-6-phosphatase



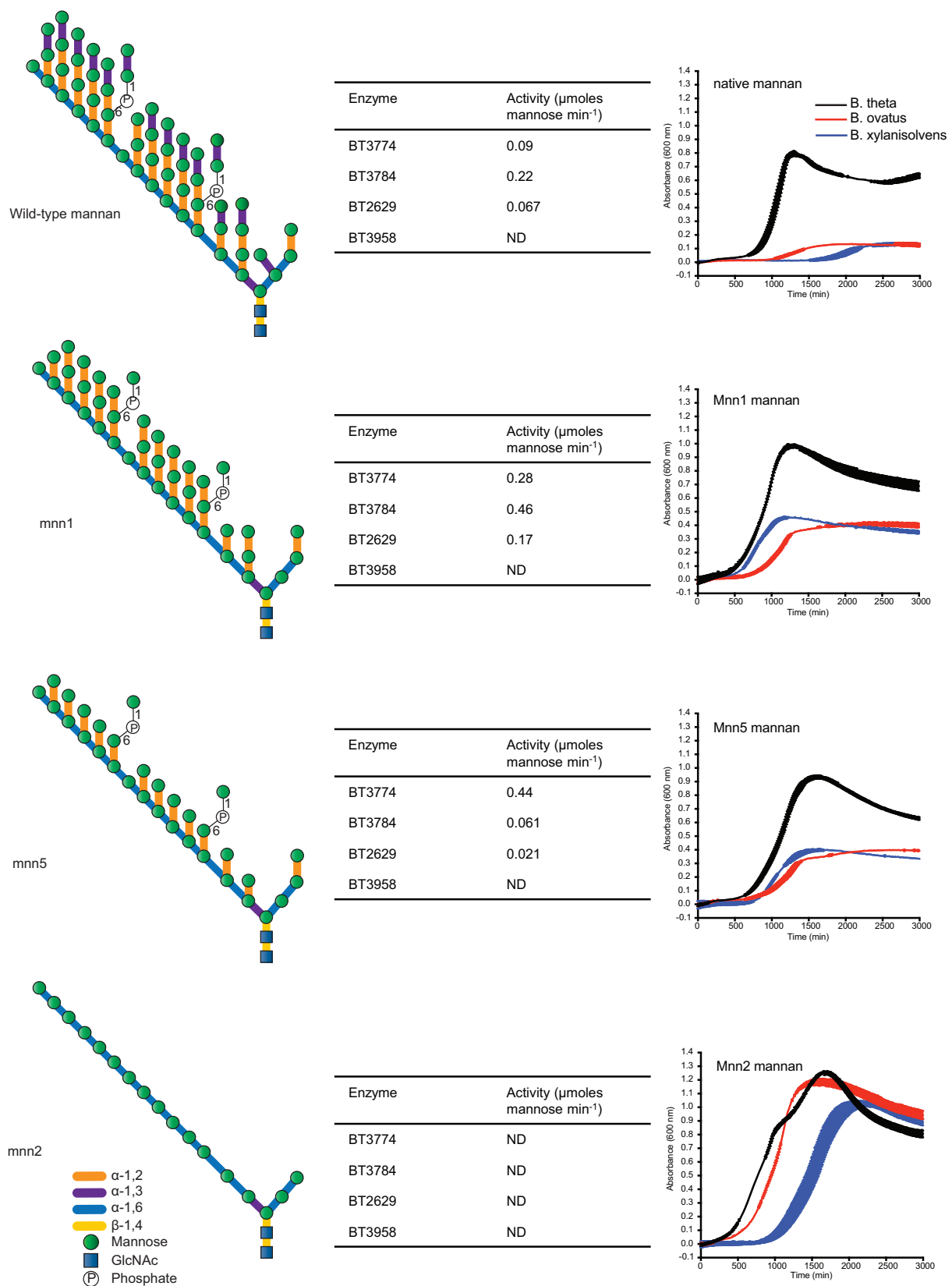
Extended Data Figure 3 | The structures of enzymes that play a key role in yeast mannan degradation. **a**, Overlay of the hydrophobic conserved residues in the predicted substrate-binding cleft of BT3792 (yellow), BT2949 (cyan) and the *Listeria* protein Lin0763 (green; PDB code 3K7X), and the predicted catalytic aspartates. **b**, Solvent representation of BT3792 in which the predicted catalytic residues, Asp258 and Asp259, are coloured green. **c**, Overlay of BT3862 (cyan) with a homologue of the enzyme from *B. xylanisolvens*, BxGH99 (green; PDB code 4UTF) in complex with Man- α -1,3-isofagomine and α -1,2-mannobiose (Man residues coloured yellow and isofagomine pink). **d**, Solvent-exposed surface of the substrate binding cleft of the BxGH99 (teal) ligand complex overlaid with BT3862 (grey). The subsites are numbered with the catalytic residues, Glu 333 and Glu 336, coloured red and the solvent exposed O2 of Man bound at the -2 subsite and O1 and O6 of the Man located at the

+2 subsite are coloured bright green. **e**, Overlay of BT3781 (green; PDB code 2P0V) with the substrate and catalytic residues of the *Clostridium perfringens* GH125 α -mannosidase CpGH125 (cyan; PDB code 3QT9), in which the ligand 6-S- α -D-mannopyranosyl-6-thio- α -D-mannopyranose (Man-S-Man) is shown in yellow. **f**, Solvent-exposed surface of BT3781 in the vicinity of the active site in which the catalytic residues (Glu 174 and Glu 439) are depicted in green. The position of Man-S-Man is based on the overlay shown in **e**. **g**, Overlay of BT3783 with the catalytic and substrate binding residues of a tyrosyl-DNA phosphodiesterase (PDB code 4GYZ) in complex with Mg^{2+} (slate-blue sphere) and phosphate (coloured orange). **h**, A region of the solvent-accessible surface of BT3783 in which the catalytic residues are coloured green. The figure was prepared using PyMOL. A detailed description of the structures of these proteins is provided in Supplementary Information section 5.0.



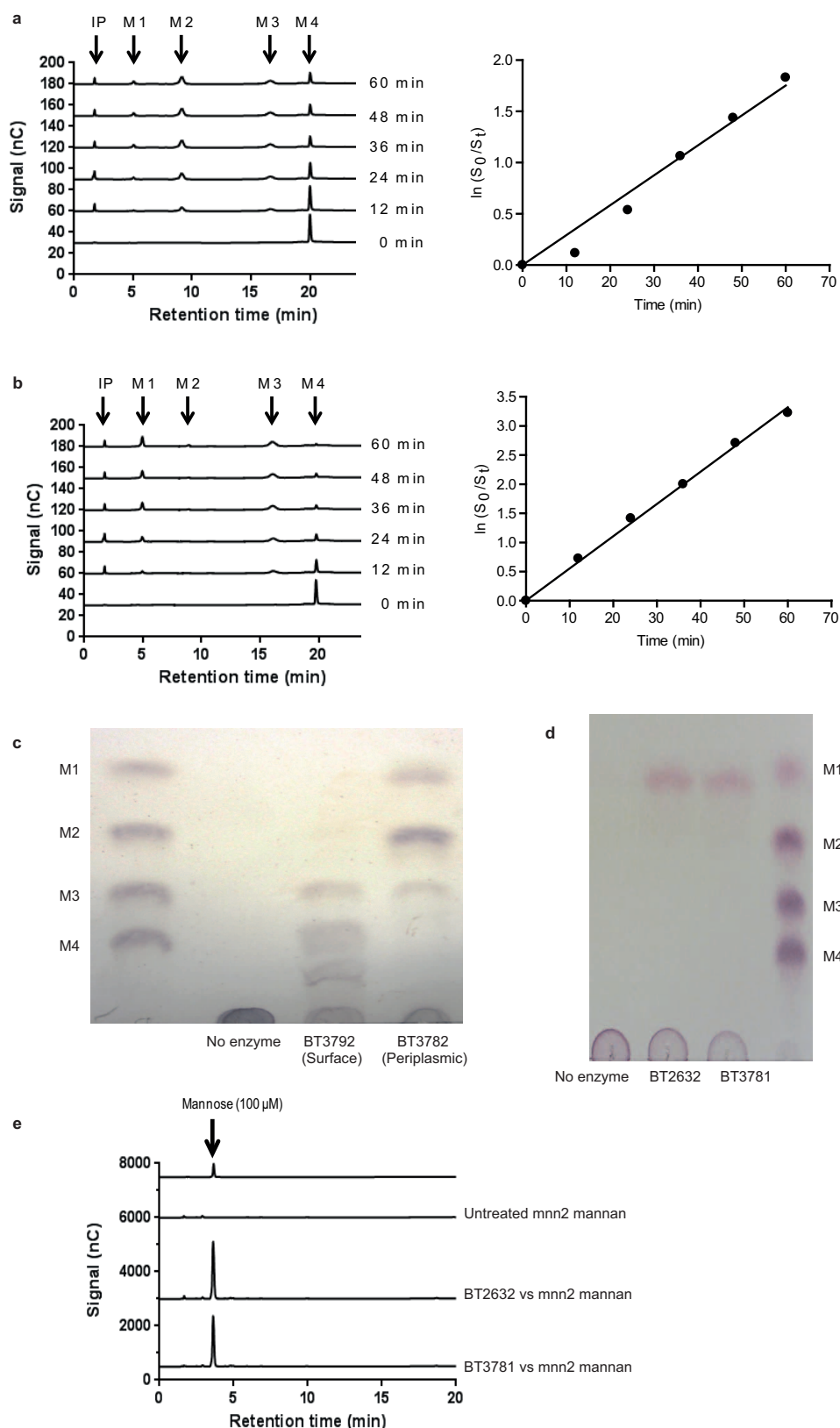
Extended Data Figure 4 | The degradation of yeast mannan by *B. thetaiotaomicron* in culture and the selected enzymes expressed by the bacterium, and the stereochemistry of the reaction catalysed by GH76 endo- α -1,6-mannanases. **a**, GH92 α -mannosidases at high concentrations (50 μ M) were incubated with yeast mannan for 5 h in the absence (labelled GH92) or in the presence (GH92/GH76) of the endo- α -1,6-mannanase BT3782. The GH92 α -mannosidases in this example were BT2199 (1), BT2130 (2) and BT3773 (3). The GH76 endo- α -1,6-mannanase only releases manno oligosaccharides in the presence of BT2199; see also Supplementary Information section 4.1. **b**, *B. thetaiotaomicron* was grown on yeast mannan or glucose. Yeast mannan was incubated with no bacterium (1), *B. thetaiotaomicron* previously cultured on yeast mannan (2) and *B. thetaiotaomicron* grown on glucose (3). The cells were incubated for 5 h at 37 °C with the polysaccharide without a nitrogen source and thus were not growing. The products released by the *B. thetaiotaomicron* cells, analysed by TLC, were mediated by the activity of enzymes presented on the surface of *B. thetaiotaomicron*, and not through the action of periplasmic mannanases and mannosidases. The black box highlights very low levels of high molecular weight manno oligosaccharides generated by the cells incubated in yeast mannan. **c**, *B. thetaiotaomicron* was cultured for up to 48 h (stationary phase) on yeast mannan. The supernatant of the culture at the time points indicated

were analysed by TLC. In all panels the samples were chromatographed with the following α -1,6-manno oligosaccharides: mannose, M1; mannobiose, M2; mannotriose, M3; mannotetraose, M4. **d**, The absorbance of the culture used in **c**. **e**, BT3792 (GH76) endo- α -1,6-mannosidase is a retaining glycoside hydrolase. Enzymatic hydrolysis of 4-nitrophenyl α -D-mannopyranosyl-1,6- α -D-mannopyranoside (S) was monitored by ^1H -NMR spectroscopy (500 MHz). The stacked spectra show the reaction progress over time. $\text{S}_{\text{H1}\alpha}$ is the anomeric proton of the reducing end mannopyranoside of the substrate, and $\text{S}_{\text{H1}'\alpha}$ is the anomeric proton of the non-reducing end mannopyranoside. The reaction proceeds with the initial formation of the product, the α -anomer of α -1,6-mannobiose (P- α , peaks $\text{P}_{\text{H1}\alpha}$ and $\text{P}_{\text{H1}'\alpha}$), which slowly mutarotates to the β -anomer (P- β , peaks $\text{P}_{\text{H1}\beta}$ and $\text{P}_{\text{H1}'\beta}$). **f**, TLC analysis of *S. cerevisiae* mannan incubated without enzyme (lane 1), BT3774 (lane 2), BT3792 (lane 3), and BT3774 and BT3792 (lane 4). M1–M4 are α -manno oligosaccharide standards numbered according to their d.p. GH76 mannanase BT3792 does not attack the backbone of *S. cerevisiae* mannan unless the side chains are first removed by the GH38 α -mannosidase BT3774, confirming that this enzyme cleaves the mannose α -1,2-linked to the mannan backbone. The data in **a**, **b** and **c** are representative of two biological replicates, while the data in **f** are representative of two technical replicates.



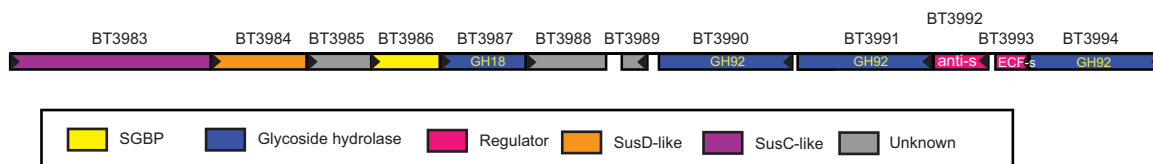
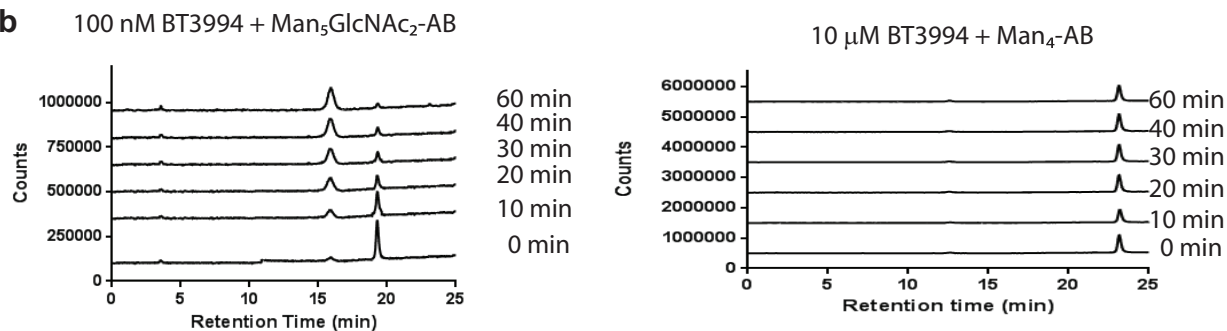
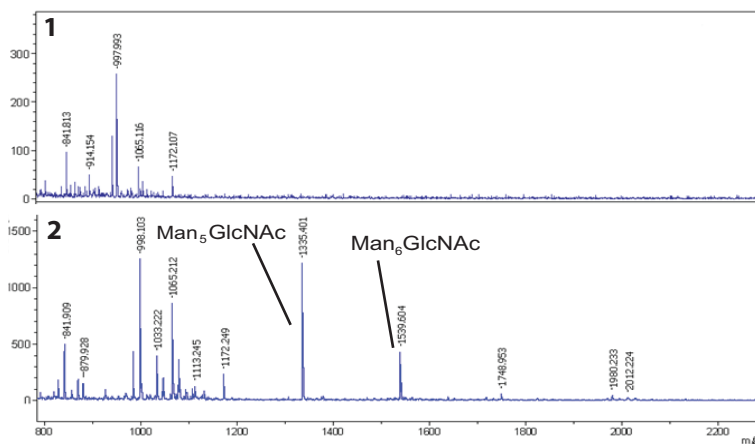
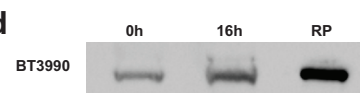
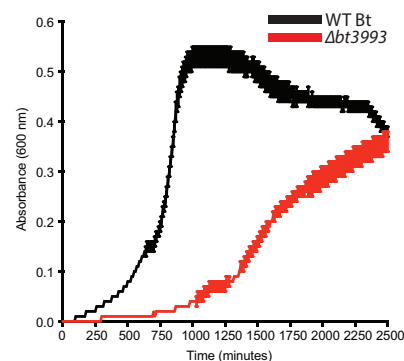
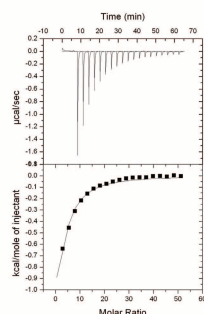
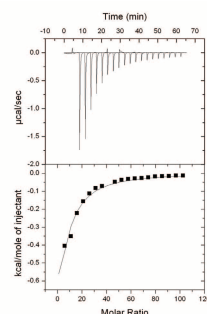
Extended Data Figure 5 | The activity of periplasmic α -mannosidases and the growth of different species of *Bacteroides* against yeast mannan. Structures of the mannans derived from wild-type and mutants of *S. cerevisiae*. The tables adjacent to the different yeast structures depict the initial rate of mannan hydrolysis by the four enzymes. The growth curves adjacent to the different mannan structures show the growth profile of *B. theta*taoimicron

(black), *B. ovatus* (red) and *B. xylanisolvens* (blue) on the glycans (each point represents the mean growth of 3 separate cultures \pm s.d.). The porcine-derived *B. xylanisolvens* strain shown here acquired MAN-PUL1 by lateral gene transfer (Extended Data Fig. 9), explaining its capacity to degrade processed mannans. Vertical error bars represent standard deviation of three separate replicates in each condition.



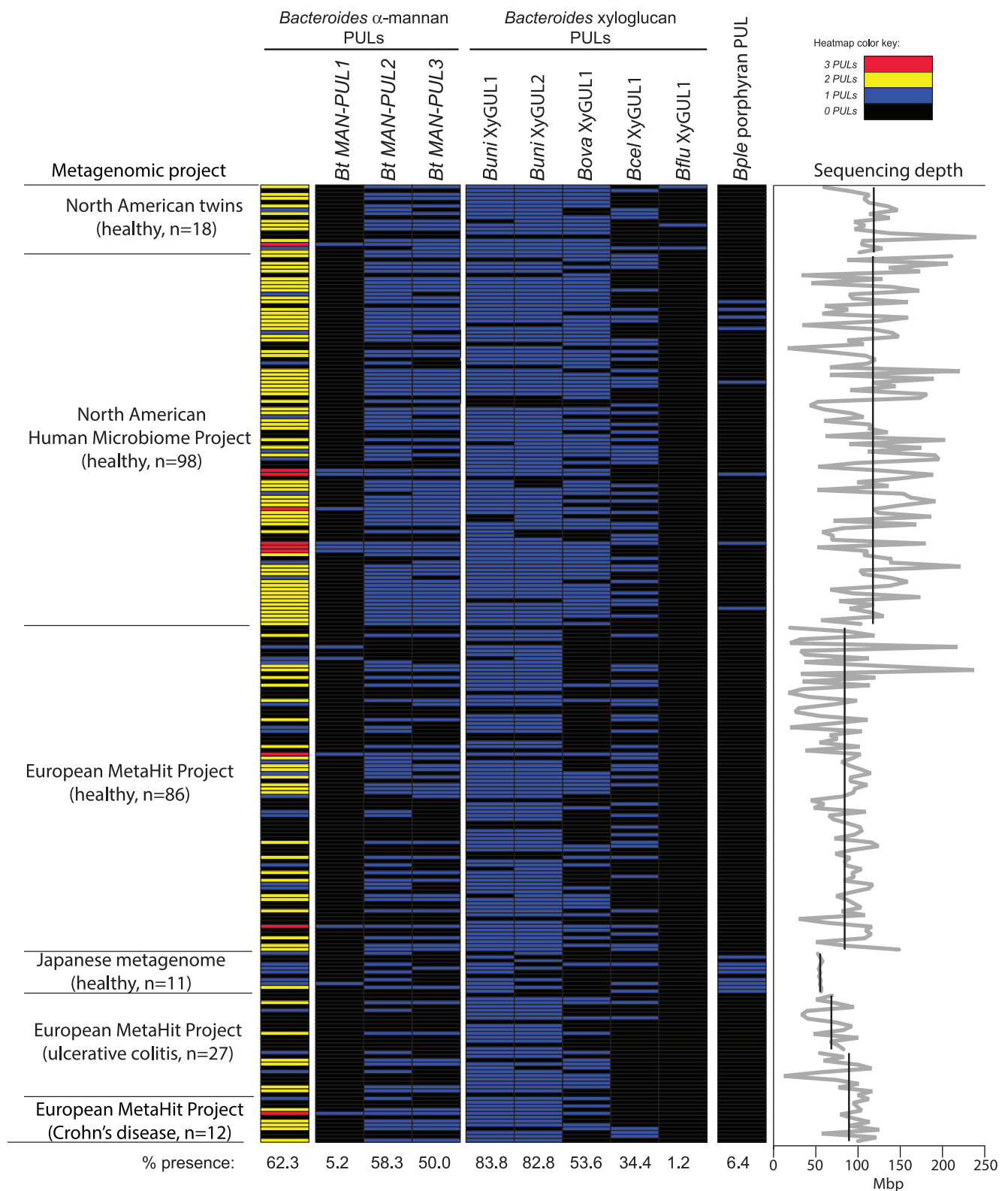
Extended Data Figure 6 | The activity of GH76 α -mannanases and GH125 α -mannosidases **a, b**, BT3792 and BT3782, respectively, were incubated with α -1,6-mannotetraose at a concentration $\ll K_M$. Substrate depletion was measured using HPAEC and the rate (right of **a** and **b**) enabled k_{cat}/K_M to be determined. **c**, BT3792 and BT3782 were incubated with unbranched yeast mannan (derived from the *S. cerevisiae* mutant MNN2). The yeast mannan at 0.1% was incubated with the two GH76 α -1,6-mannanases for 1 h at 37 °C in

50 mM sodium phosphate buffer, pH 7.0. The limit products were analysed by TLC. α -1,6-Mannooligosaccharides are identified by their degree of polymerization (M1, mannose; M2, mannobiose; M3, mannotriose; M4, mannotetraose); IP, injection peak. **d, e**, BT2632 and BT3781 at 100 nM were incubated with 1 mg ml⁻¹ of the debranched mannan for 1 h in the buffer described above. TLC analysis (**d**) and HPAEC traces (**e**) of the reactions are shown. The data in **c** and **d** are representative of two technical replicates.

a**b****c****d****g****e****f**

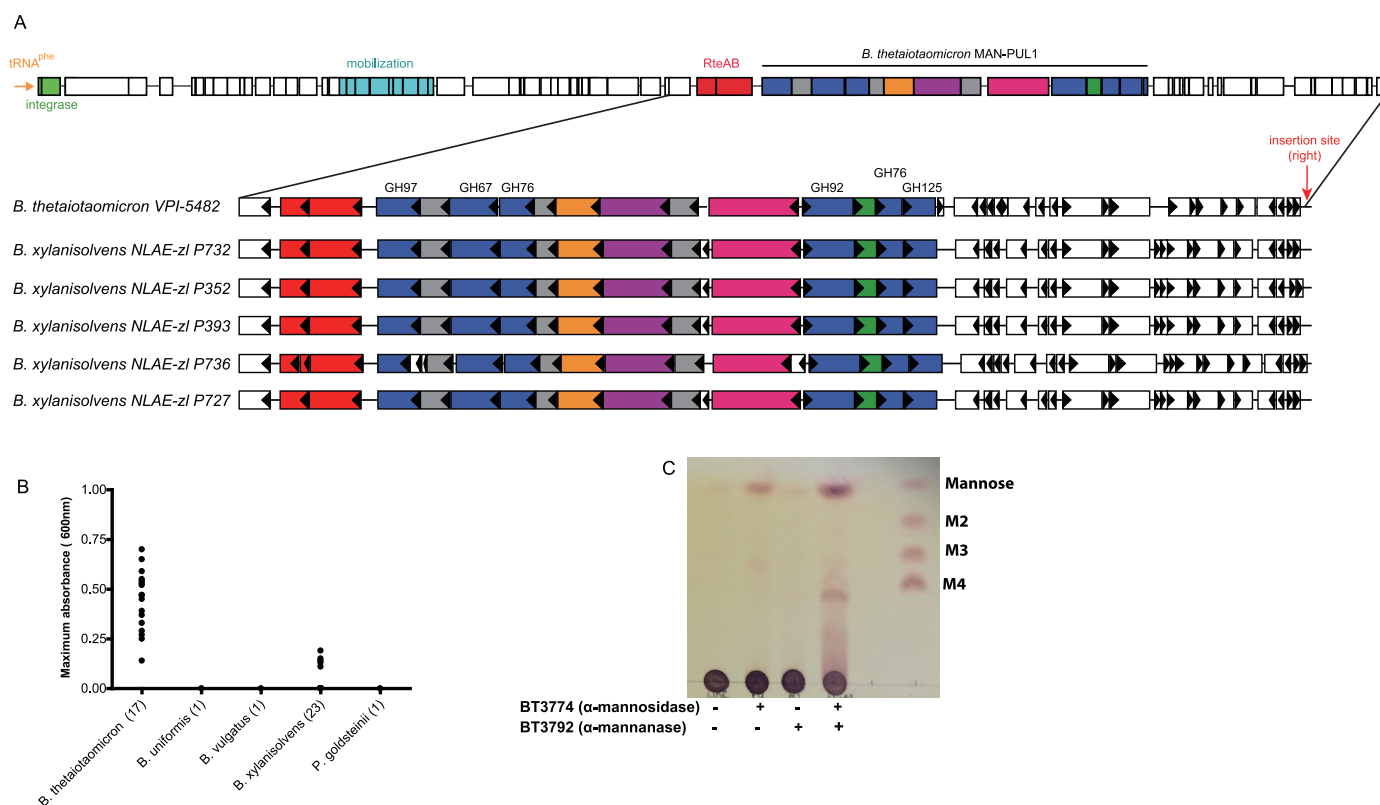
Extended Data Figure 7 | HMNG deconstruction by *B. thetaiotaomicron*.
a, Structure of the HMNG PUL. Genes drawn to scale with their orientation indicated. Genes encoding known or predicted functionalities are colour-coded and, where appropriate, are also annotated according to their CAZy glycoside hydrolase (GH) family number. SGBP represents a surface glycan binding protein. **b**, BT3994 was incubated with α -1,6-mannotetraose (Man_4) or the high mannose *N*-glycan $\text{Man}_5\text{GlcNAc}_2$, with both oligosaccharides labelled with 2-aminobenzamide (AB). At the indicated time points aliquots were removed and analysed by HPAEC using a fluorescence detection system. While $\text{Man}_5\text{GlcNAc}_2$ -AB was hydrolysed by BT3994, the enzyme was not active against Man_4 -AB. **c**, Chicken ovalbumin was incubated with buffer (1) or 1 μM of BT3987 (2) in 20 mM Na-HEPES buffer, pH 7.5, for 5 h at 37 °C, and the soluble material was permethylated and analysed by MALDI-TOF mass spectrometry. The high mannose *N*-glycans released are labelled. **d**, Western blot of *B. thetaiotaomicron* cells cultured on yeast mannan that were untreated

with proteinase K (0 h) or incubated with 2 mg ml^{-1} proteinase K for 16 h (16 h). The lane labelled RP contained a purified recombinant form of BT3990. The blots were probed with antibodies against BT3990. The data in **d** are representative of two biological replicates. **e**, **f**, Representative isothermal calorimetry titrations for BT3984 titrated with Gal- β 1,4-GlcNAc (LacNAc; 25 mM) (**e**), and for BT3986 titrated with mannose (50 mM) (**f**). The top half of each panel shows the raw isothermal calorimetry titration heats; the bottom half, the integrated peak areas fitted using a single binding model by MicroCal Origin software. ITC was carried out in 50 mM Na-HEPES, pH 7.5 at 25 °C. The affinities and thermodynamic parameters of binding are shown in Supplementary Table 5. **g**, Growth profile of wild-type *B. thetaiotaomicron* (WT Bt; black) and the mutant *Δbt3993* (red), which lacks the extra-cellular factor σ regulator gene of HMNG-PUL, cultured on $\text{Man}_5\text{GlcNAc}_2$ (each curve shows the mean \pm s.d. of 3 separate cultures).



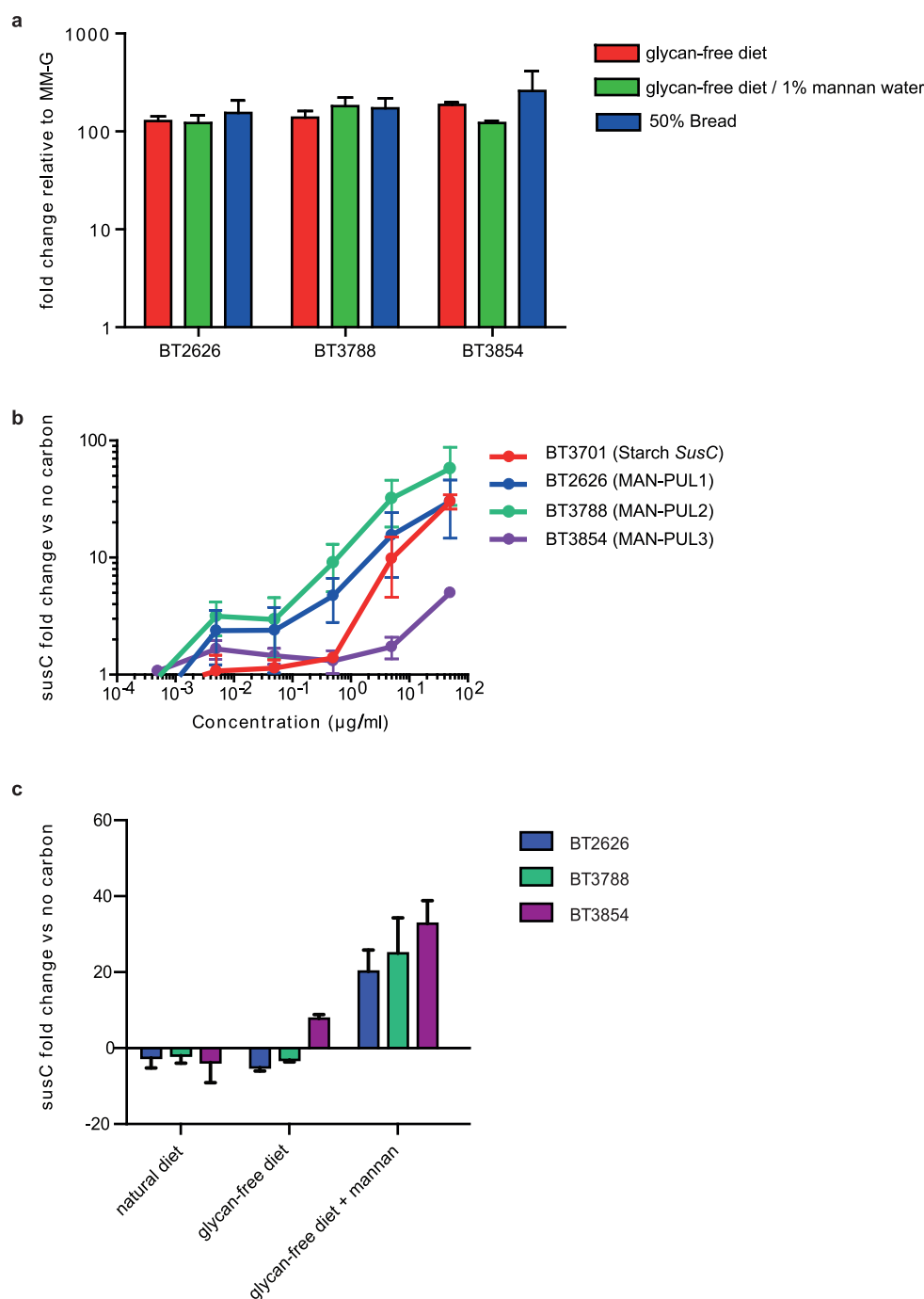
Extended Data Figure 8 | Metagenomic analysis of the occurrence of the yeast mannan PULs in humans. Abundance of *Bacteroides* mannan PULs in humans from a survey of metagenomic sequencing data from a total of 250 adult human samples (211 healthy, 27 ulcerative colitis, 12 Crohn's disease; see Methods for references). Data sets were individually queried by BLAST using either each entire mannan PUL (PULs 2,3) or a sub-fragment that was trimmed to eliminate cross-detection of other species genomes beyond *B. thetaiotaomicron* and porcine *B. xylanisolvens* (PUL1; see Methods for additional search details). Each horizontal line represents the presence of a hit

in a single individual. The leftmost column summarizes the total mannan PUL content in each person (annotated according to the colour key in the upper right corner). The mannan PUL frequency across all 250 samples is shown at the bottom for each condition and is compared to the frequency of several other PULs implicated in xyloglucan and porphyrin utilization. Graph at far right illustrates the variation in sequencing depth for each sample/study; black lines show the average depth in megabase pairs (Mbp) for each study; the light grey line shows the depth for each individual sample.



Extended Data Figure 9 | Presence of a conjugative transposon (cTn) that contains MAN-PUL1 in the genomes of porcine *B. xylanisolvens* strains and the mannan presented to these organisms. **a**, Shown across the top is a schematic of a cTn that has been integrated into the 3' end of a $tRNA^{Phe}$ gene in *B. thetaiotaomicron* strain VPI-5482. Integration is mediated by a 22 bp direct repeat sequence that is contained in $tRNA^{Phe}$ and repeated again at the other side of the cTn (right insertion site). The location of *B. thetaiotaomicron* MAN-PUL1 is denoted within the larger cTn element using a colour scheme identical to Fig. 1a. The lower panel shows an expanded view of the MAN-PUL1 locus in five sequenced strains of *B. xylanisolvens* from the faeces of pigs fed a diet enriched with distillers grains that were fermented with yeast. A nearly identical copy (both by amino acid homology and syntenic organization) of this genomic region is present in *B. thetaiotaomicron* and the

porcine *B. xylanisolvens* strains. Although the draft genomes of the *B. xylanisolvens* strains contains gaps in all five assemblies at the left side of the MAN-PUL1, the right side insertion site was resolved in all genomes, suggesting that the *B. xylanisolvens* loci were also transferred by lateral transfer at some point in the history of these strains. **b**, Forty-three different strains from five *Bacteroidetes* isolated from animal guts (each indicated with a solid circle) were inoculated into minimal media containing *S. cerevisiae* mannan as the sole carbon source. The growth of the cultures was measured over 48 h by recording the optical density at 600 nm. **c**, TLC analysis of the products generated by incubating BT3774 and BT3792 with α -mannan extracted from the distillers grain fed to the pigs from which the *B. xylanisolvens* were isolated. The data are representative of two technical replicates.



Extended Data Figure 10 | In vivo and in vitro expression of the mannan PULs. **a**, Level of *susC*-like transcripts derived from MAN-PUL1 (BT2626), MAN-PUL2 (BT3788) and MAN-PUL3 (BT3854) from *B. thetaiotaomicron* in monocolonized gnotobiotic mice fed a glycan-free diet deficient in *B. thetaiotaomicron*-digestible glycans (red), the same diet with added yeast mannan (1% w/v in drinking water) as the only usable polysaccharide (green), and a diet containing 50% bread (blue). The levels of the *susC* transcripts were quantified relative to the same mRNA species in *B. thetaiotaomicron* cultured *in vitro* on glucose minimal medium (MM-G). Note that in all cases, expression of MAN-PUL genes is equally high *in vivo*. **b**, Transcription of the same mannan *susC*-like genes in response to increasing concentrations of yeast

mannan in the media after 30 min exposure. The prototypic *susC* (BT3701) involved in starch metabolism is shown for comparison. **c**, An identical exposure experiment to that shown in **b**, except that glucose-grown *B. thetaiotaomicron* cells were exposed to aqueous extracts of the cereal grain diet (natural diet) fed to mice before the experiment, or the digestible glycan-free control diet (glycan-free diet) used as a base for all feeding treatments. Exposure was conducted for 30 min to determine if any diet extract contained contaminating levels of mannan that could be detected by *B. thetaiotaomicron* cells; inclusion of purified mannan (5 mg ml^{-1}) in addition to the glycan-free diet served as positive controls. In all panels, the results represent the mean of 3 biological replicates and error bars represent s.d.

Glutathione activates virulence gene expression of an intracellular pathogen

Michelle L. Reniere¹, Aaron T. Whiteley², Keri L. Hamilton³, Sonya M. John¹, Peter Lauer⁴, Richard G. Brennan³ & Daniel A. Portnoy^{1,5}

Intracellular pathogens are responsible for much of the world-wide morbidity and mortality due to infectious diseases. To colonize their hosts successfully, pathogens must sense their environment and regulate virulence gene expression appropriately. Accordingly, on entry into mammalian cells, the facultative intracellular bacterial pathogen *Listeria monocytogenes* remodels its transcriptional program by activating the master virulence regulator PrfA. Here we show that bacterial and host-derived glutathione are required to activate PrfA. In this study a genetic selection led to the identification of a bacterial mutant in glutathione synthase that exhibited reduced virulence gene expression and was attenuated 150-fold in mice. Genome sequencing of suppressor mutants that arose spontaneously *in vivo* revealed a single nucleotide change in *prfA* that locks the protein in the active conformation (PrfA*) and completely bypassed the requirement for glutathione during infection. Biochemical and genetic studies support a model in which glutathione-dependent PrfA activation is mediated by allosteric binding of glutathione to PrfA. Whereas glutathione and other low-molecular-weight thiols have important roles in redox homeostasis in all forms of life, here we demonstrate that glutathione represents a critical signalling molecule that activates the virulence of an intracellular pathogen.

Listeria monocytogenes is a Gram-positive pathogen of animals and humans that cycles between a saprophytic lifestyle and an intracellular pathogen that escapes from a vacuole and grows in the cytosol of host cells¹. The intracellular lifecycle of *L. monocytogenes* has been well characterized and is entirely dependent on the transcription factor PrfA (refs 2, 3). PrfA directly regulates the transcription of nine virulence factors and is therefore referred to as the master virulence regulator in *L. monocytogenes*⁴. In keeping with its central role in pathogenesis, *L. monocytogenes* strains lacking *prfA* are completely avirulent^{1,3}. PrfA is a member of the cAMP receptor protein (Crp) family of transcription factors, which are characterized by their allosteric regulation via small-molecule activators. In *L. monocytogenes*, PrfA is exclusively activated in the cytosol of host cells, leading to the assumption that the activating cofactor for PrfA is specific to this compartment. However, even after decades of study, the biochemical mechanism by which PrfA detects the intracellular environment is not well understood. The goal of this study was to identify how *L. monocytogenes* recognizes and responds to its intracellular niche of the mammalian cell cytosol.

Genetic selection in macrophages

We devised a genetic selection to isolate bacterial mutants unable to activate virulence genes during intracellular growth. Our strategy took advantage of a *L. monocytogenes* vaccine strain designed to die *in vivo* (P.L. *et al.*, manuscript in preparation). Specifically, *loxP* sites were inserted into the *L. monocytogenes* chromosome flanking the origin of replication (*ori*, Fig. 1a). Into this background a codon-optimized *cre* recombinase gene was inserted under the control of the *actA* promoter, which is the most exquisitely regulated PrfA-dependent virulence gene in *L. monocytogenes* and is specifically activated in the host cytosol^{2,3,5,6}. The resulting strain grew like wild type *in vitro* (Fig. 1b) where *actA* expression is very low^{4,5}. However, on cytosolic access, Cre-mediated recombination of the *loxP* sites resulted in excision of the *ori*, preventing

bacterial replication (Fig. 1c). A transposon library was then generated in this 'suicide' strain background. Bone-marrow-derived macrophages (BMDMs) were infected with the library of transposon mutants and the surviving bacteria were recovered.

We identified more than 16 independent insertions in a *L. monocytogenes* gene, previously identified as encoding a bifunctional glutathione synthase (*gshF*)⁷, that rescued the death of the 'suicide' strain *in vivo* (Fig. 1c, d). Glutathione is a tripeptide low-molecular-weight (LMW) thiol present in all eukaryotes that contain mitochondria and nearly all Gram-negative bacteria⁸. *L. monocytogenes* is one of the few Gram-positive bacteria that synthesize glutathione, whereas many utilize alternative LMW thiols such as bacillithiol and mycothiol^{9,10}. Glutathione was not required for Cre/*lox* recombination when *cre* was expressed from a constitutive promoter (data not shown), leading to the hypothesis that glutathione was required specifically for activation of the *actA* promoter.

Glutathione is required for virulence

To determine the role of *gshF* in *L. monocytogenes*, an in-frame deletion strain was generated by allelic exchange ($\Delta gshF$). Consistent with published work⁷, the *gshF*-deficient strain was moderately more sensitive to oxidative stress *in vitro* (Extended Data Fig. 1). However, $\Delta gshF$ did not suffer a general loss of fitness, as it exhibited no measurable growth defect *in vitro* (Fig. 2a) or in BMDMs (Fig. 2b). As expected based on the criteria of the genetic selection, the $\Delta gshF$ mutant expressed lower levels of ActA in cells (Fig. 2c), formed very small plaques in tissue culture assays that measure cell-to-cell spread (Fig. 2d), and was greater than 2-logs less virulent in mice (Fig. 2e). Complementation of $\Delta gshF$ with its native promoter ($\Delta gshF + gshF$) restored wild-type ActA abundance, wild-type plaque size, and virulence (Fig. 2c–e). Since all mammalian cells have high intracellular levels of glutathione¹¹, we assessed the potential contribution of host glutathione using buthionine sulfoximine

¹Department of Molecular and Cell Biology, University of California, Berkeley, California 94720, USA. ²Graduate Group in Infectious Diseases and Immunity, School of Public Health, University of California, Berkeley, California 94720, USA. ³Department of Biochemistry, Duke University School of Medicine, Durham, North Carolina 27710, USA. ⁴Aduro BioTech, Inc. Berkeley, California 94710, USA. ⁵School of Public Health, University of California, Berkeley, California 94720, USA.

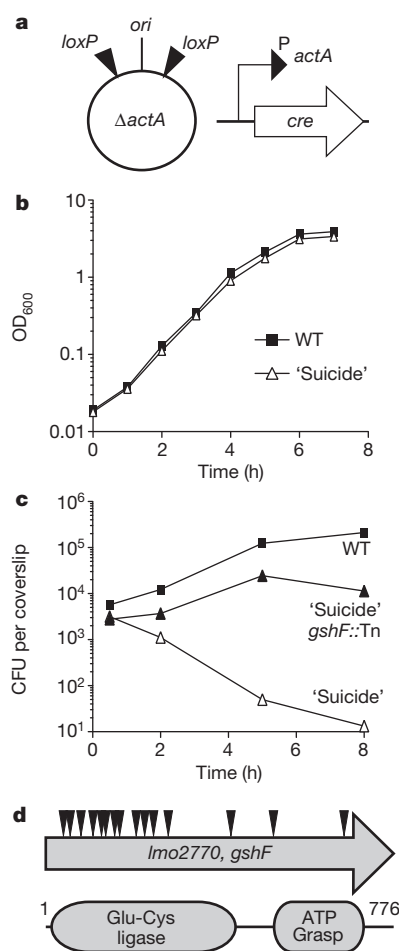


Figure 1 | Forward genetic selection to identify factors required for virulence gene activation during infection. **a**, Schematic of the 'suicide' strain used for genetic selection. See text for description. **b**, Broth growth curve. Data are representative of three independent experiments. OD_{600} , optical density at 600 nm. **c**, BMDM growth curve. Data are a combination of three independent experiments. **d**, Schematic of transposon insertions identified in $gshF$ and the conserved protein domains.

$\Delta gshF$ mutant was due to imported host glutathione and also established that the phenotypes observed for $\Delta gshF$ were due to a lack of glutathione and not absence of the GshF protein.

Isolation of suppressor mutations *in vivo*

To elucidate the role of glutathione during infection we sought to isolate suppressor mutations. The selective pressure of the host environment was used to select for compensatory mutations in the $\Delta gshF$ background that restored virulence to identify functionally interacting genes and/or pathways. Since previous work identified $gshF::Tn$ mutants as hypo-haemolytic¹³, we screened for hyper-haemolytic colonies from the liver homogenates of infected animals on blood agar plates. Two hyper-haemolytic colonies were isolated and genome sequencing identified a single nucleotide polymorphism (SNP) common to both strains and absent from the $\Delta gshF$ parental strain. The SNP encoded a PrfA G145S mutation, which is the most commonly found spontaneous PrfA* allele¹⁴, so called because of its structural similarity to well-characterized Crp* mutants that are constitutively active in the absence of cofactor¹⁵. The PrfA G145S allele rescued ActA expression and virulence of $\Delta gshF$, confirming the function of this mutation identified by our *in vivo* suppressor analysis (Fig. 3a–c). This was not specific to $actA$, as transcript levels of two other PrfA-dependent genes were also restored by the PrfA G145S mutation (Extended Data Fig. 3). Furthermore, two other previously identified PrfA* alleles¹⁶ also rescued the plaque defect of $\Delta gshF$ (Fig. 3d), indicating that constitutively activating PrfA completely bypassed the requirement for glutathione during infection. Importantly, these data highlighted that $\Delta gshF$ was not attenuated during infection due to a general loss of fitness, but rather, due to a dysregulation of virulence genes.

PrfA binds glutathione allosterically

In addition to its role in maintaining redox homeostasis, glutathione can be covalently bonded to protein thiols as a post-translational modification,

(BSO). BSO depletes total cellular glutathione levels >98%¹², but had no effect on bacterial growth during infection (Extended Data Fig. 2). Whereas wild-type *L. monocytogenes* was unaffected, the $\Delta gshF$ mutant failed to synthesize detectable ActA in the BSO-treated cells (Fig. 2f). These results demonstrated that the remaining ActA expression in the

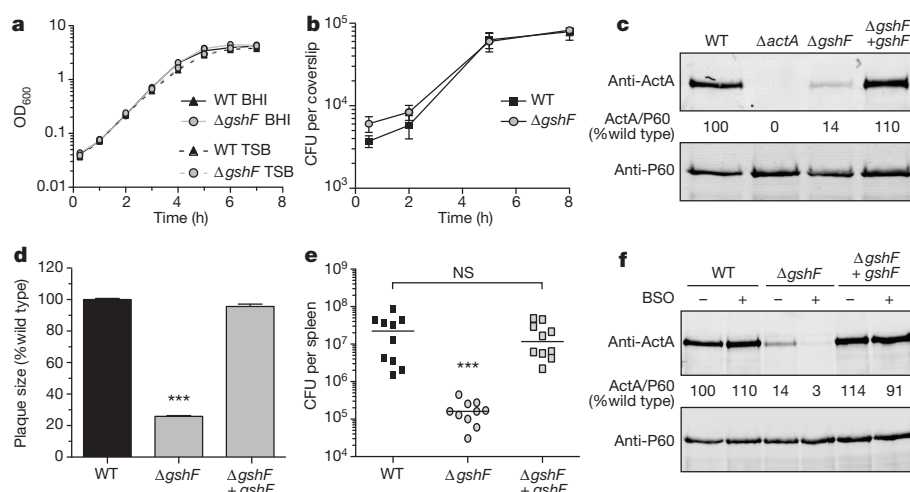


Figure 2 | *Listeria monocytogenes* $\Delta gshF$ is attenuated *in vivo*. **a**, Broth growth curve in brain heart infusion (BHI) or tryptic soy broth (TSB). Data are representative of three independent experiments. **b**, BMDM growth curve. Mean \pm standard error of the mean (s.e.m.) for three independent experiments is shown. **c**, Representative immunoblot of infected BMDMs. Numbers are the mean of four independent experiments and indicate ActA normalized to P60, as a per cent of wild type. **d**, Plaque size. Mean \pm s.e.m. for three

independent experiments is shown. **e**, CD-1 mice were infected intravenously and analysed as described in Methods. Data are a combination of two independent experiments, $n = 10$ mice per strain. The median of each group is represented as a horizontal line. **f**, Representative immunoblot of infected BMDMs. Quantification is as described in **c**. In all panels P values were calculated using Student's t -test; *** $P < 0.001$; NS denotes $P > 0.05$.

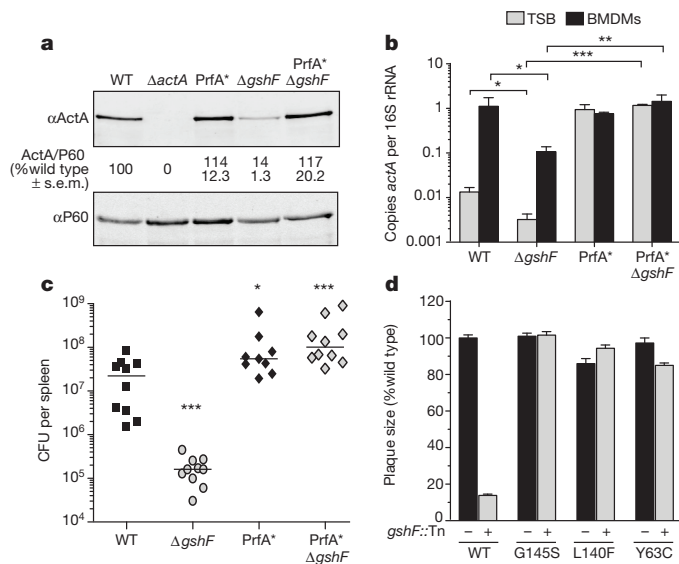


Figure 3 | PrfA* bypasses the requirement for glutathione during infection. **a**, Representative immunoblot of infected BMDMs. Quantification is as described in Fig. 2. **b**, Quantitative reverse transcription polymerase chain reaction (RT-PCR) of *actA* transcript abundance. Mean \pm s.e.m. for three independent experiments is shown. **c**, Mice were infected as described in Fig. 2. Data are a combination of two independent experiments, $n = 10$ per strain. The median of each group is represented as a horizontal line. **d**, Plaque size. Mean \pm s.e.m. for three independent experiments is shown. In all panels asterisks denote a significant difference compared to wild type, unless otherwise indicated, as determined by Student's *t*-test; * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$.

a process referred to as S-glutathionylation¹⁷. To determine whether glutathionylation of PrfA is required for its activation, we engineered a PrfA protein in which all four cysteine residues were mutated to alanine (referred to as PrfA(C38A/C144A/C205A/C229A) or PrfA(C/A)₄). Recombinant PrfA(C/A)₄ bound DNA with an affinity similar to wild-type PrfA *in vitro* (Table 1 and Extended Data Fig. 4), establishing that these mutations do not disturb the overall structural integrity of the protein. However, the cysteine residues were found to contribute to DNA binding, as demonstrated by the 25-fold lower affinity of oxidized wild-type PrfA as compared to reduced PrfA (Table 1). Although PrfA(C/A)₄ bound DNA *in vitro*, it was less abundant than wild type when expressed from the native locus on the chromosome of *L. monocytogenes* (Extended Data Fig. 5). Since PrfA is auto-regulated¹⁸, these data suggested that PrfA(C/A)₄ was less active *in vivo*. Indeed, the PrfA(C/A)₄ strain synthesized less ActA than the wild-type strain during BMDM infection (Fig. 4a) and was 30-fold less virulent in mice (Fig. 4b). Together, these results suggested that the cysteine residues of PrfA were

Table 1 | DNA-binding and glutathione-binding affinities of PrfA

	DNA-binding affinity ($K_d \pm$ s.e.m.)	
	Phly (nM)	PactA (nM)
Wild type (oxidized)	888.5 \pm 140.3	ND
Wild type (reduced)	34.2 \pm 4.9	96.4 \pm 7.3
PrfA(C/A) ₄	32.8 \pm 5.5	124.9 \pm 26.3
PrfA*	40.8 \pm 3.3	45.4 \pm 3.2
	Glutathione-binding affinity ($K_d \pm$ s.e.m.)	
	GSH (mM)	GSSG (mM)
Wild type	4.37 \pm 1.2	NBD
PrfA(C/A) ₄	4.74 \pm 1.5	NBD

DNA-binding affinity for the *hly* promoter (Phly) and the *actA* promoter (PactA), as measured by fluorescence anisotropy, and glutathione-binding affinity, as measured by bio-layer interferometry. The affinity of oxidized PrfA to PactA was not determined (ND). DNA-binding affinities of PrfA(C/A)₄ and PrfA* were unaffected by oxidation. For oxidized glutathione (GSSG) no measurable binding was detected (NBD).

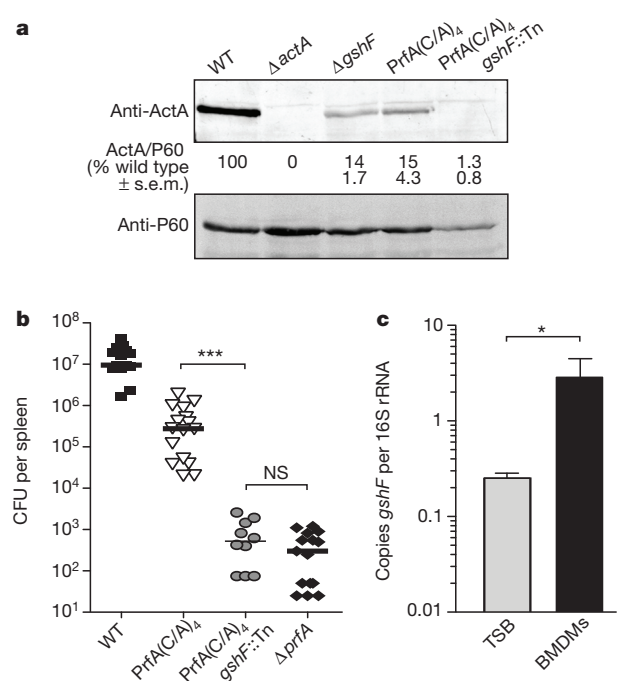


Figure 4 | Glutathione-dependent PrfA activation is mediated by allosteric binding, not glutathionylation. **a**, Representative immunoblot of infected BMDMs. Quantification is as described in Fig. 2. **b**, Mice were infected as described in Fig. 2. Data are a combination of at least two independent experiments, $n = 10$ or 16 per strain. The median of each group is represented as a horizontal line. All strains were significantly different from wild type ($P < 0.001$). **c**, Quantitative RT-PCR of *gshF* transcript abundance. Mean \pm s.e.m. for three independent experiments are shown. In all panels asterisks denote a significant difference compared to wild type, unless otherwise indicated, as determined by Student's *t*-test; * $P < 0.05$, *** $P < 0.001$; NS denotes $P > 0.05$.

dispensable for DNA binding *in vitro*—as the mutant lacking all cysteine residues (PrfA(C/A)₄) bound DNA with an affinity similar that of the wild type (Table 1)—but were required for activity *in vivo*. If glutathionylation of PrfA was important for its activity, then deleting *gshF* in the PrfA(C/A)₄ background would have no effect. Remarkably, combining the PrfA(C/A)₄ and *gshF*::Tn mutations resulted in a strain that was defective for intracellular growth (Extended Data Fig. 6) and completely avirulent in mice (>4 -log attenuation, Fig. 4b).

Although we cannot definitively exclude a role for glutathionylation of PrfA *in vivo*, covalent modification of the protein thiols was not sufficient for PrfA activation, as the PrfA(C/A)₄ mutant was only modestly attenuated during infection (Fig. 4b). Therefore, we hypothesized that glutathione may be allosterically binding PrfA, analogous to the interaction of cAMP binding to Crp¹⁹. The binding affinity of glutathione for PrfA was measured via bio-layer interferometry. A direct and specific interaction with reduced glutathione (GSH) was detected for both recombinant wild-type and PrfA(C/A)₄ with binding affinities of 4.4 ± 1.2 and 4.7 ± 1.5 mM, respectively, whereas no measurable interaction was found between either protein and oxidized glutathione (GSSG, Table 1). Although the affinity for reduced glutathione appears to be relatively low, it is well within biologically relevant concentrations of this LMW thiol, as both prokaryotes and eukaryotes have intracellular concentrations of 0.1–10 mM glutathione^{8,11}. This binding affinity would also allow PrfA to be sensitive to varying concentrations of glutathione, rather than being a simple ON–OFF switch. In support of the hypothesis that glutathione may activate PrfA *in vivo*, *gshF* was transcriptionally upregulated tenfold during infection (Fig. 4c). These data demonstrate that reduced glutathione non-covalently binds PrfA and support the model that glutathione is the activating cofactor for PrfA.

Discussion

The results of this study clearly demonstrate that both bacterial and host-synthesized glutathione contribute to expression of *L. monocytogenes* virulence factors via allosteric binding of the master virulence regulator PrfA. Unlike Crp, PrfA does not require allosteric activation to bind DNA *in vitro* (Table 1). Indeed, the DNA-binding affinity of PrfA was unchanged in the presence of glutathione (data not shown). In this regard, PrfA is similar to the phylogenetically more closely related Crp family member NtcA from cyanobacteria, which also binds DNA in the absence of its cofactor^{20,21}. Together, our results suggest a model whereby PrfA activation is a two-step process requiring reduced protein thiols for initial DNA binding and allosteric binding of glutathione to PrfA for transcriptional activation (see model in Extended Data Fig. 7). Indeed, eliminating both of these steps, as in the PrfA(C/A)₄ *gshF*::Tn mutant, results in a strain that is as attenuated as a Δ prfA mutant (Fig. 4b).

Glutathione is present in the cytosol of all host cells so it is perhaps not surprising that intracellular pathogens import it, as is the case with *Francisella tularensis*²². What is surprising is that *L. monocytogenes* imports glutathione from the host and also synthesizes it. The results of this study suggest that *L. monocytogenes* uses glutathione concentration to regulate its biphasic lifestyle and the switch from saprophyte to pathogen. This may be a reflection of the broad host range of this pathogen and the fact that glutathione is ubiquitous in host cells, making it a reliable signal of the *L. monocytogenes* cytosolic niche. Perhaps other LMW thiols, such as coenzyme A, mycothiol and bacillithiol have similar activating roles in virulence gene expression in other pathogens.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 25 June; accepted 3 November 2014.

- Xayarath, B. & Freitag, N. E. Optimizing the balance between host and environmental survival skills: lessons learned from *Listeria monocytogenes*. *Future Microbiol.* **7**, 839–852 (2012).
- Freitag, N. E., Port, G. C. & Miner, M. D. *Listeria monocytogenes*—from saprophyte to intracellular pathogen. *Nature Rev. Microbiol.* **7**, 623–628 (2009).
- Chakraborty, T. *et al.* Coordinate regulation of virulence genes in *Listeria monocytogenes* requires the product of the *prfA* gene. *J. Bacteriol.* **174**, 568–574 (1992).
- de las Heras, A., Cain, R. J., Bielecka, M. K. & Vázquez-Boland, J. A. Regulation of *Listeria* virulence: PrfA master and commander. *Curr. Opin. Microbiol.* **14**, 118–127 (2011).
- Moors, M. A., Levitt, B., Youngman, P. & Portnoy, D. A. Expression of listeriolysin O and ActA by intracellular and extracellular *Listeria monocytogenes*. *Infect. Immun.* **67**, 131–139 (1999).
- Shetron-Rama, L. M., Marquis, H., Bouwer, H. G. A. & Freitag, N. E. Intracellular induction of *Listeria monocytogenes* *actA* expression. *Infect. Immun.* **70**, 1087–1096 (2002).
- Gopal, S. *et al.* A multidomain fusion protein in *Listeria monocytogenes* catalyzes the two primary activities for glutathione biosynthesis. *J. Bacteriol.* **187**, 3839–3847 (2005).
- Masip, L., Veeravalli, K. & Georgiou, G. The many faces of glutathione in bacteria. *Antioxid. Redox Signal.* **8**, 753–762 (2006).
- Newton, G. L. *et al.* Distribution of thiols in microorganisms: mycothiol is a major thiol in most actinomycetes. *J. Bacteriol.* **178**, 1990–1995 (1996).
- Newton, G. L. *et al.* Bacillithiol is an antioxidant thiol produced in Bacilli. *Nature Chem. Biol.* **5**, 625–627 (2009).
- Meister, A. & Anderson, M. E. Glutathione. *Annu. Rev. Biochem.* **52**, 711–760 (1983).
- Rouzer, C. A., Scott, W. A., Griffith, O. W., Hamill, A. L. & Cohn, Z. A. Depletion of glutathione selectively inhibits synthesis of leukotriene C by macrophages. *Proc. Natl Acad. Sci. USA* **78**, 2532–2536 (1981).
- Zemansky, J. *et al.* Development of a mariner-based transposon and identification of *Listeria monocytogenes* determinants, including the peptidyl-prolyl isomerase PrsA2, that contribute to its hemolytic phenotype. *J. Bacteriol.* **191**, 3950–3964 (2009).
- Ripio, M. T., Domínguez-Bernal, G., Lara, M., Suárez, M. & Vázquez-Boland, J. A. A Gly145Ser substitution in the transcriptional activator PrfA causes constitutive overexpression of virulence factors in *Listeria monocytogenes*. *J. Bacteriol.* **179**, 1533–1540 (1997).
- Eiting, M., Hagelüken, G., Schubert, W.-D. & Heinz, D. W. The mutation G145S in PrfA, a key virulence regulator of *Listeria monocytogenes*, increases DNA-binding affinity by stabilizing the HTH motif. *Mol. Microbiol.* **56**, 433–446 (2005).
- Miner, M. D., Port, G. C. & Freitag, N. E. Functional impact of mutational activation on the *Listeria monocytogenes* central virulence regulator PrfA. *Microbiology* **154**, 3579–3589 (2008).
- Dalle-Donne, I., Rossi, R., Colombo, G., Giustarini, D. & Milzani, A. Protein S-glutathionylation: a regulatory device from bacteria to humans. *Trends Biochem. Sci.* **34**, 85–96 (2009).
- Mengaud, J. *et al.* Pleiotropic control of *Listeria monocytogenes* virulence factors by a gene that is autoregulated. *Mol. Microbiol.* **5**, 2273–2283 (1991).
- Kolb, A., Busby, S., Buc, H., Garges, S. & Adhya, S. Transcriptional regulation by cAMP and its receptor protein. *Annu. Rev. Biochem.* **62**, 749–797 (1993).
- Valladares, A., Flores, E. & Herrero, A. Transcription activation by NtcA and 2-oxoglutarate of three genes involved in heterocyst differentiation in the cyanobacterium *Anabaena* sp. strain PCC 7120. *J. Bacteriol.* **190**, 6126–6133 (2008).
- Körner, H., Sofia, H. J. & Zumft, W. G. Phylogeny of the bacterial superfamily of Crp-Fnr transcription regulators: exploiting the metabolic spectrum by controlling alternative gene programs. *FEMS Microbiol. Rev.* **27**, 559–592 (2003).
- Alkhuder, K., Meibom, K. L., Dubail, I., Dupuis, M. & Charbit, A. Glutathione provides a source of cysteine essential for intracellular multiplication of *Francisella tularensis*. *PLoS Pathog.* **5**, e1000284 (2009).

Acknowledgements We thank N. Freitag for providing strains and P. Hwang (UCSF Biosensor Core Facility) for technical support and advice regarding bio-layer interferometry. This work used the Vincent J. Coates Genomics Sequencing Laboratory at UC Berkeley, supported by NIH S10 Instrumentation grants S10RR029668 and S10RR027303 and the UCSF Funding Shared Equipment Award. This work was supported by National Institutes of Health grants 1P01 AI63302 and 1R01 AI27655 to D.A.P.; M.L.R. is supported by F32AI104247; A.T.W. is supported by the NSF GRFP DGE 1106400; K.L.H. is supported by F32GM008487.

Author Contributions M.L.R., A.T.W., K.L.H. and S.M.J. performed the experiments; P.L. engineered the 'suicide' strain; M.L.R., A.T.W., R.G.B. and D.A.P. designed the study; M.L.R. and D.A.P. wrote the paper. All authors discussed the results and commented on the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare competing financial interests: details are available in the online version of the paper. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to D.A.P. (portnoy@berkeley.edu).

METHODS

Bacterial strains and cell culture. The *L. monocytogenes* strains used were all in the 10403S background (Extended Data Table 1). Bacteria were cultured in brain heart infusion (BHI) or tryptic soy broth (TSB), which contains less than 0.5 µM glutathione⁹. All media were from Becton Dickinson (New Jersey). For broth growth curves, overnight cultures were diluted 1:100 and optical density at 600 nm (OD₆₀₀) was measured at each time point using a spectrophotometer. *gshF* (*lmo2770*) was deleted by allelic exchange using the temperature-sensitive plasmid pKSV7 (refs 23, 24). The Δ *gshF* complemented strain was generated by inserting a C-terminal 6×His-tagged copy of *gshF* with its native promoter into the integration vector pPL2 (ref. 25).

Murine L2 fibroblasts were passaged in Dulbecco modified Eagle medium with high glucose (DMEM, Gibco/Invitrogen) supplemented with 1% sodium pyruvate, 1% L-glutamine, and 10% fetal bovine serum (FBS, GemCell) at 37 °C with 5% CO₂. Bone-marrow-derived macrophages (BMDMs) were cultured in DMEM supplemented with 1% sodium pyruvate, 1% L-glutamine, 20% FBS and 10% 3T3-MCSF supernatant at 37 °C with 5% CO₂.

Transposon library generation and genetic selection. A transposon library was generated in the 'suicide' strain using *himar1 mariner* transposon mutagenesis, as previously described¹³. BMDMs were then infected with this library of transposon mutants. Cells were collected at various time points after infection, lysed, and surviving bacteria were plated on BHI agar. Individual colonies were then isolated and used to infect BMDMs to confirm the phenotype. To identify the transposon insertion site, colony PCR was performed using primers listed in Extended Data Table 1. A large percentage of the transposon insertion sites were found in the *actA* promoter, *cre*, and each *loxP* site. The fact that we identified multiple transposon insertions in each *loxP* site, which are less than 40 nucleotides, indicates that the genetic selection approached saturation.

We next screened these mutants by plaque assay and mutants with a plaque size <90% of wild type were included. Finally, intracellular growth curves were performed to ensure that the mutants had a defect only in *actA* expression and not in intracellular growth. Greater than 16 independent insertions were identified in *lmo2770* (*gshF*), making it by far the most over-represented hit in the selection.

Intracellular growth curves. BMDMs were harvested as previously described²⁶ and 3 × 10⁶ cells were plated in 60 mm non-TC-treated Petri dishes. Cells were infected with a multiplicity of infection (MOI) of 0.1 and growth curves were performed as described previously²⁷.

Quantitative RT-PCR of bacterial transcripts. For transcript analysis in broth, bacteria were grown overnight in TSB and subcultured 1:100 into 25 ml TSB. Bacteria were harvested at an OD₆₀₀ = 1.0. For transcript analysis during infection, BMDMs were plated at a density of 3 × 10⁷ cells in 150 mm TC-treated dishes and infected with an MOI of 10. One hour post-infection the cells were washed and media containing gentamicin (50 µg ml⁻¹) was added. Four hours post-infection the cells were washed and lysed in 0.1% NP-40 containing RNAlprotect Bacteria Reagent (Qiagen). Bacteria were harvested by centrifugation.

After harvesting bacteria from either broth or BMDMs they were lysed in phenol:chloroform containing 1% SDS by vortexing with 0.1 mm diameter silica/zirconium beads (BioSpec Products Inc.). Nucleic acids were precipitated from the aqueous fraction overnight at -80 °C in ethanol containing 150 mM sodium acetate (pH 5.2). Precipitated nucleic acids were washed with ethanol and treated with TURBO DNase per manufacturer's specifications (Life Technologies Corporation). RNA was again precipitated overnight and then washed in ethanol. RT-PCR was performed with iScript Reverse Transcriptase (Bio-Rad) and quantitative PCR (qPCR) of resulting cDNA was performed with KAPA SYBR Fast (Kapa Biosystems). Primers used for qPCR are listed in Extended Data Table 1.

Plaque assay. Plaque assays in L2 murine fibroblasts were performed as previously described²⁸. Briefly, bacterial cultures were grown overnight at 30 °C, then washed and diluted 1:10 in sterile PBS. Six-well dishes containing 1.2 × 10⁶ L2 cells per well were infected with *L. monocytogenes* for 1 h, then washed and overlaid with 3 ml of media containing 0.7% agarose and gentamicin (10 µg ml⁻¹) to prevent extracellular growth. After 3 days at 37 °C, an overlay containing gentamicin and neutral red dye (Sigma) was added and stained overnight. The plates were then scanned and analysed with ImageJ software²⁹.

Immunoblots of infected BMDMs. BMDMs were plated in 12-well dishes at a density of 10⁶ cells per well and infected with an MOI of 10. One hour post-infection the cells were washed and media containing gentamicin (50 µg ml⁻¹) was added. Where indicated, 2 mM buthionine sulfoximine (Santa Cruz Biotechnology) was added to cells 16 h before infection and included throughout the infection. Four hours post-infection the cells were washed and harvested in LDS buffer containing 5% BME. The samples were then boiled and separated by 10% SDS-PAGE. A rabbit polyclonal antibody against the N terminus of ActA³⁰ and a mouse monoclonal antibody against P60 (Adipogen) were each used at a dilution of 1:5,000. P60 is a constitutively expressed bacterial protein³¹ used as a loading control.

Virulence experiments. Six-to-eight-week-old female CD-1 mice (The Jackson Laboratory) were infected intravenously with 1 × 10⁵ colony-forming units (CFU). Forty-eight hours post-infection the mice were euthanized and spleens and livers were harvested, homogenized and plated for enumeration of bacterial burdens. Sample size was chosen based on standards within the field^{32,33}. No statistical methods were used to predetermine sample size. Samples were not blinded or randomized. All statistical tests allowed for unequal variance between groups, that is, two-tailed heteroscedastic *t*-test. All animal work was done in accordance with university regulations. Protocols were reviewed and approved by the Animal Care and Use Committee at the University of California, Berkeley (MAUP# R235-0813B).

In vivo suppressor analysis. Female CD-1 mice were infected intravenously with 1 × 10⁷ CFU of Δ *gshF* for 72 h and the livers were harvested, homogenized and inoculated into broth. New mice were then infected with 1 × 10⁶ CFU of the bacteria from the liver homogenates. Seventy-two hours post-infection the livers were harvested, homogenized and plated on blood agar plates. Two hyper-haemolytic colonies were observed and were chosen for further analysis.

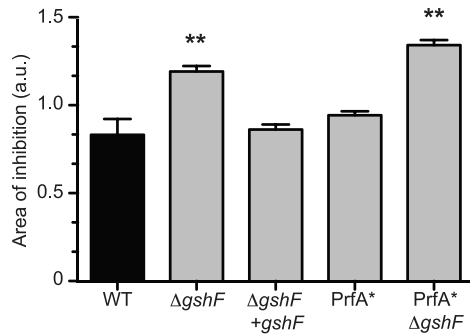
Genome sequencing. *L. monocytogenes* genomic DNA was extracted (MasterPure Kit, Epicentre) and sequenced by Illumina 50SR (library preparation and sequencing performed by UC Berkeley QB3 Genomic Sequencing Laboratory). Sequencing data were aligned to the 10403S reference genome and SNP/InDel/structural variation was determined (CLC Genomics Workbench, CLC bio) for the Δ *gshF* parent strain and the two hyper-haemolytic suppressor mutants.

Protein purification and binding analyses. Recombinant PrfA was purified from *Escherichia coli* BL21(DE3) as previously described³⁴. Glutathione binding to the wild-type or PrfA(C/A)₄ protein was measured by bio-layer interferometry on an Octet RED 384 instrument (Pall ForteBio). The buffer used was PBS, pH 7.3 containing 2 mM tris(2-carboxyethyl)phosphine (TCEP) when appropriate. Samples or buffer were dispensed into 384-well microtitre plates at a volume of 100 µl per well. Operating temperature was maintained at 26 °C with 1,000 rpm rotary agitation. Ni-NTA biosensor tips (Pall ForteBio) were pre-soaked for 10 min with buffer to establish a baseline before protein immobilization. 6×His-tagged proteins diluted in PBS, pH 7.3 were immobilized onto the biosensors for 8 min at a concentration of 35 µg ml⁻¹. The immobilization level attained was 7–8 nm. Binding association of the glutathione with biosensor tips was monitored for 30 s, and subsequent dissociation in buffer was monitored for 30 s. Glutathione was tested at concentrations of 0.5, 1, 1.5, 2, 3, 4, 5 mM. Reduced glutathione (GSH) was diluted in buffer containing TCEP, while oxidized glutathione (GSSG) was diluted in PBS only. To control for background, association and dissociation of GSH/GSSG was measured with biosensor tips loaded with buffer only and biosensor tips loaded with protein were tested for binding in buffer with or without TCEP. The apparent affinities of glutathione and PrfA were calculated from equilibrium measurements and, when appropriate, global fits of the *k*_{on} and *k*_{off} values, yielding similar values. Data are mean and s.e.m. of experiments from four independent protein preparations.

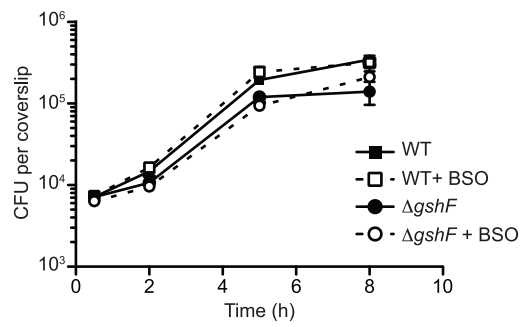
Fluorescence polarization. A fluorescence-polarization-based DNA-binding assay was used to determine the affinities of PrfA, PrfA* and the PrfA(C/A)₄ mutant for the *Phly* and *PactA* promoters. The sequences of the top strands of *Phly* and *PactA* used in this study are in Extended Data Table 1. The oligodeoxynucleotides were purchased from IDT (Coralville, Iowa) with a fluorescein label covalently attached to the 5' end. DNA binding was measured in PBS buffer (11.8 mM Na⁺/K⁺ phosphate, 2.7 mM KCl, 137 mM NaCl) at 25 °C using 5 nM fluoresceinated target dsDNA, and 1 µg poly(dI-dC) as a nonspecific DNA competitor. In some experiments 1 mM TCEP was included to maintain a reducing environment. PrfA was titrated into the DNA until saturation as denoted by no further change in the millipolarization (mP = units of polarization × 10⁻³). The fluoresceinated DNA was excited at 490 nm and its parallel and perpendicular emission intensities measured at 530 nm and converted to units of mP using a Beacon 2000 Variable Temperature Fluorescence Polarization System. Data were plotted and analysed with the following equation: mP = {(mP_{bound} - mP_{free})[protein]}/K_d + [protein] + mP_{free}, where mP is the millipolarization measured at a given protein concentration, mP_{free} is the initial millipolarization of free fluorescein-labelled DNA, mP_{bound} is the maximum millipolarization of specifically bound DNA, and [protein] is the protein concentration. The generated hyperbolic curves are fit by nonlinear least-squares regression analysis, assuming a bimolecular model such that the K_d values represent the protein concentration at half-maximal ligand binding and plotted by using the graphing program, Kaleidograph. The K_d values are expressed in terms of PrfA dimer binding.

23. Smith, K. & Youngman, P. Use of a new integrational vector to investigate compartment-specific expression of the *Bacillus subtilis* *spoIIIM* gene. *Biochimie* **74**, 705–711 (1992).
24. Camilli, A., Tilney, L. G. & Portnoy, D. A. Dual roles of *plcA* in *Listeria monocytogenes* pathogenesis. *Mol. Microbiol.* **8**, 143–157 (1993).

25. Lauer, P., Chow, M. Y. N., Loessner, M. J., Portnoy, D. A. & Calendar, R. Construction, characterization, and use of two *Listeria monocytogenes* site-specific phage integration vectors. *J. Bacteriol.* **184**, 4177–4186 (2002).
26. Sauer, J.-D. *et al.* The *N*-ethyl-*N*-nitrosourea-induced Goldenticket mouse mutant reveals an essential function of Sting in the *in vivo* interferon response to *Listeria monocytogenes* and cyclic dinucleotides. *Infect. Immun.* **79**, 688–694 (2011).
27. Portnoy, D. A., Jacks, P. S. & Hinrichs, D. J. Role of hemolysin for the intracellular growth of *Listeria monocytogenes*. *J. Exp. Med.* **167**, 1459–1471 (1988).
28. Sun, A. N., Camilli, A. & Portnoy, D. A. Isolation of *Listeria monocytogenes* small-plaque mutants defective for intracellular growth and cell-to-cell spread. *Infect. Immun.* **58**, 3770–3778 (1990).
29. Schneider, C. A., Rasband, W. S. & Eliceiri, K. W. NIH Image to ImageJ: 25 years of image analysis. *Nature Methods* **9**, 671–675 (2012).
30. Lauer, P. *et al.* Constitutive activation of the PrfA regulon enhances the potency of vaccines based on live-attenuated and killed but metabolically active *Listeria monocytogenes* strains. *Infect. Immun.* **76**, 3742–3753 (2008).
31. Köhler, S., Bubert, A., Vogel, M. & Goebel, W. Expression of the *iap* gene coding for protein p60 of *Listeria monocytogenes* is controlled on the posttranscriptional level. *J. Bacteriol.* **173**, 4668–4674 (1991).
32. Sauer, J.-D. *et al.* *Listeria monocytogenes* triggers AIM2-mediated pyroptosis upon infrequent bacteriolysis in the macrophage cytosol. *Cell Host Microbe* **7**, 412–419 (2010).
33. Melton-Witt, J. A., McKay, S. L. & Portnoy, D. A. Development of a single-gene, signature-tag-based approach in combination with alanine mutagenesis to identify listeriolysin O residues critical for the *in vivo* survival of *Listeria monocytogenes*. *Infect. Immun.* **80**, 2221–2230 (2012).
34. Böckmann, R., Dickneite, C., Middendorf, B., Goebel, W. & Sokolovic, Z. Specific binding of the *Listeria monocytogenes* transcriptional regulator PrfA to target sequences requires additional factor(s) and is influenced by iron. *Mol. Microbiol.* **22**, 643–653 (1996).
35. Bishop, D. K. & Hinrichs, D. J. Adoptive transfer of immunity to *Listeria monocytogenes*. The influence of *in vitro* stimulation on lymphocyte subset requirements. *J. Immunol.* **139**, 2005–2009 (1987).
36. Skoble, J., Portnoy, D. A. & Welch, M. D. Three regions within ActA promote Arp2/3 complex-mediated actin nucleation and *Listeria monocytogenes* motility. *J. Cell Biol.* **150**, 527–538 (2000).
37. Cheng, L. W. & Portnoy, D. A. *Drosophila* S2 cells: an alternative infection model for *Listeria monocytogenes*. *Cell. Microbiol.* **5**, 875–885 (2003).

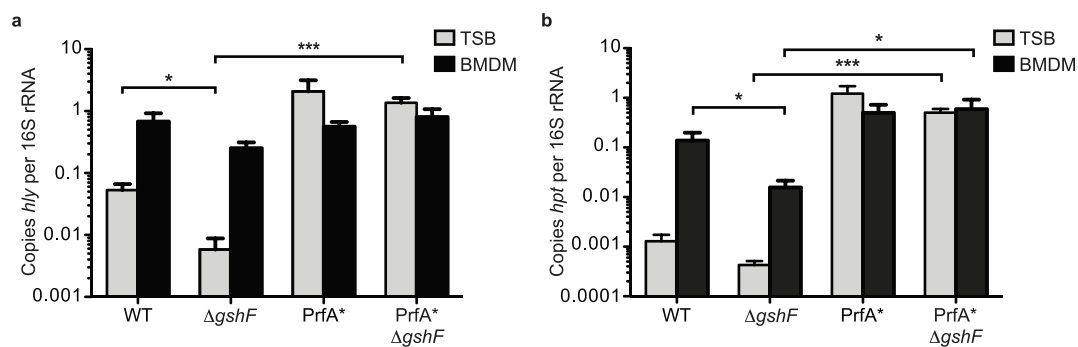


Extended Data Figure 1 | *Listeria monocytogenes* $\Delta gshF$ is sensitive to hydrogen peroxide. Bacteria were grown overnight in TSB and then inoculated into top agar and spread on tryptic soy agar plates. Sterile disks soaked in 10 μ l of 15% H_2O_2 (Thermo Fisher Scientific) were placed on the agar and incubated overnight. Plates were then scanned and the area of inhibition was measured (in arbitrary units) using ImageJ software (<http://rsbweb.nih.gov/ij>). The mean \pm s.e.m. of four independent experiments is shown. *P* values were calculated using Student's *t*-test; ***P* < 0.01. a.u., arbitrary units.



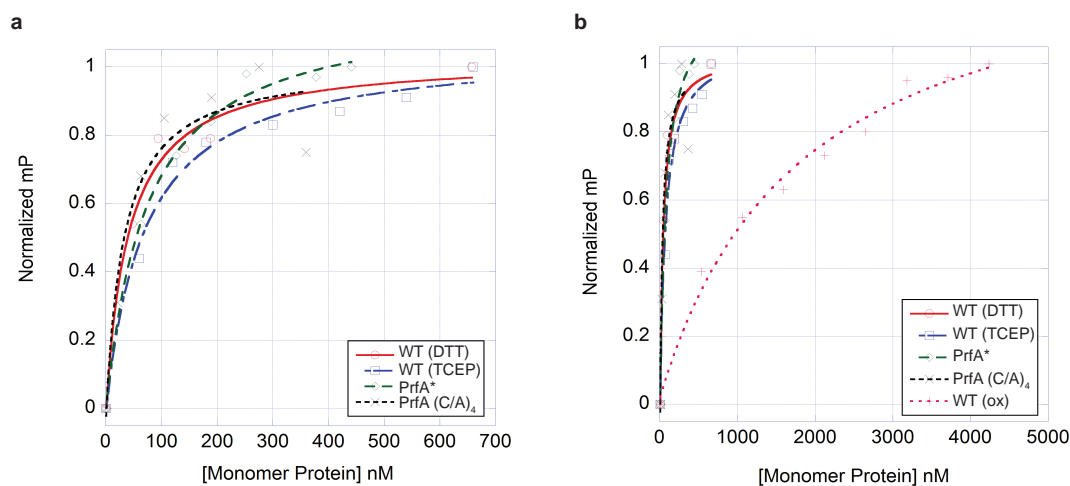
Extended Data Figure 2 | BSO does not affect *L. monocytogenes* growth.

BMDM growth curve in which cells were untreated or treated with 2 mM BSO for 16 h before infection and throughout the infection. The mean \pm s.e.m. of three independent experiments is shown.



Extended Data Figure 3 | The effect of $\Delta gshF$ is not specific to *actA* regulation. Quantitative RT-PCR of *hly* (a) or *hpt* (b) transcript levels. Bacteria were harvested from TSB at mid-log (grey bars) or 4 h post-infection of

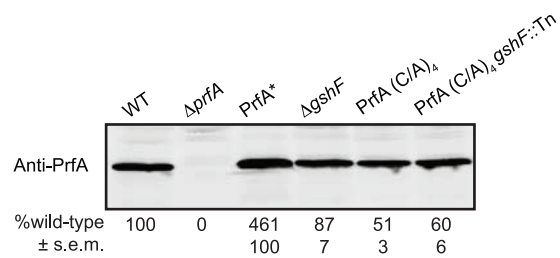
BMDMs (black bars). Mean \pm s.e.m. of three independent experiments is shown. *P* values were calculated using Student's *t*-test. **P* < 0.05; ****P* < 0.001.



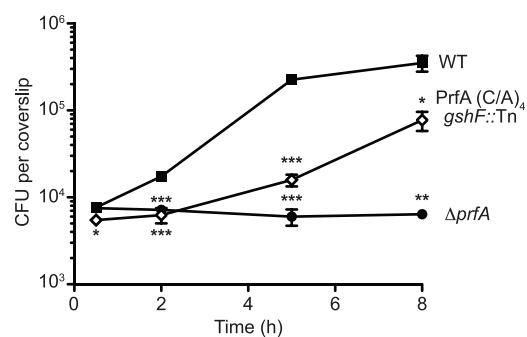
Extended Data Figure 4 | Fluorescence polarization binding isotherms.

a, Representative binding isotherms of wild-type PrfA plus DTT (circles), wild-type PrfA plus TCEP (squares), PrfA* (diamonds), and PrfA(C/A)₄ (crosses), to the PrfA box of *Phly*. **b**, Representative binding isotherms of wild-type PrfA plus DTT (circles), wild-type PrfA plus TCEP (squares), PrfA* (diamonds), PrfA(C/A)₄ (crosses), and oxidized wild-type PrfA (plus symbols),

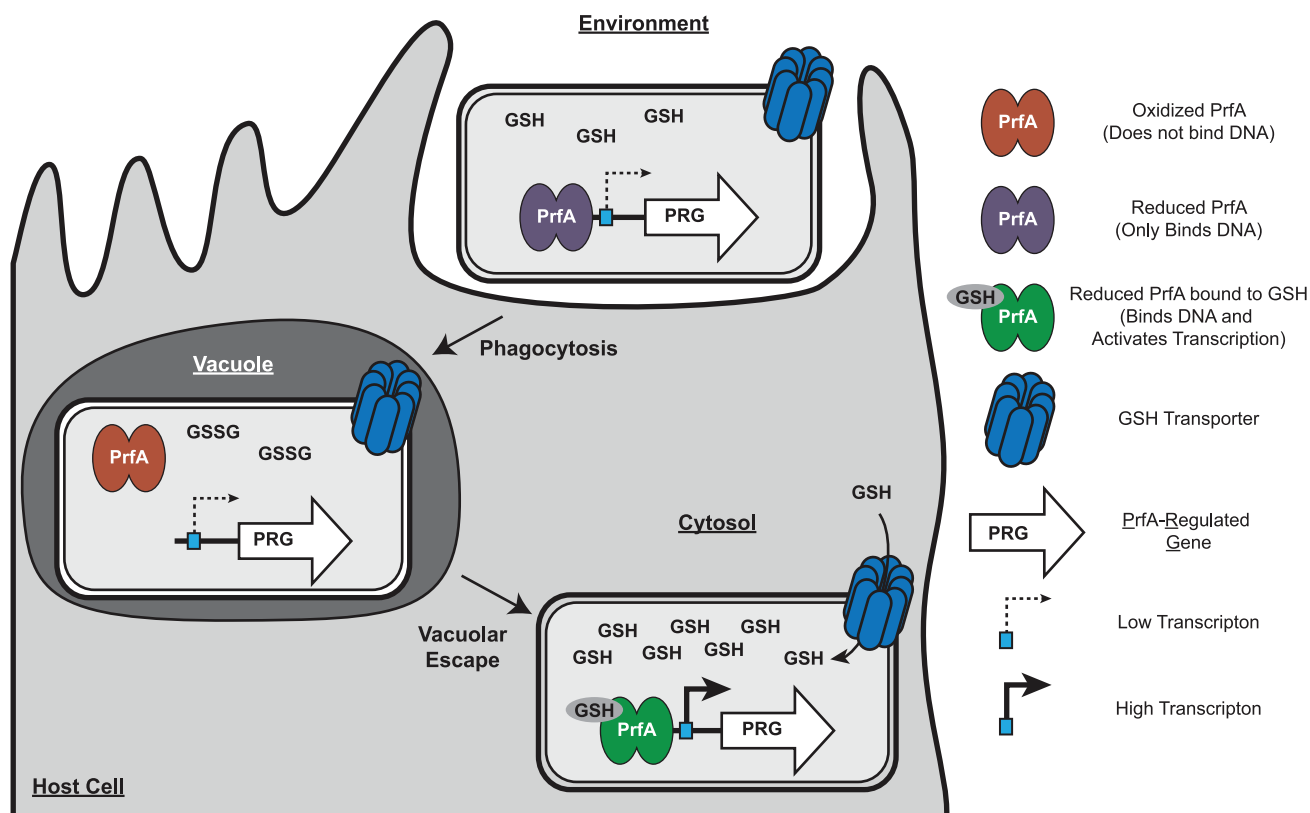
to the PrfA box of *Phly*. This plot underscores the very poor binding of oxidized wild-type PrfA to the PrfA box. In both panels the units of millipolarization (mP, y axis) have been normalized to allow the presentation of all binding isotherms on one graph. The protein concentration is shown in terms of protomer on the x axis.



Extended Data Figure 5 | PrfA(C/A)₄ expression in *L. monocytogenes* grown in broth. Immunoblot of PrfA in *L. monocytogenes* lysates harvested at early exponential phase in BHI. Mean ± s.e.m. of four independent experiments is shown.



Extended Data Figure 6 | The *PrfA(C/A)₄ gshF::Tn* mutant exhibits a significant intracellular growth defect. The mean \pm s.e.m. of four independent experiments is shown. *P* values were calculated using Student's *t*-test; **P* < 0.05; ***P* < 0.01; ****P* < 0.001.



Extended Data Figure 7 | Model of glutathione-dependent PrfA activation.

The process of infection or intercellular spread requires that *L. monocytogenes* inhabit an oxidizing vacuole, which may contain both reactive oxygen and nitrogen species. Upon oxidation, glutathione dimerizes to GSSG, which we have demonstrated does not bind PrfA. In addition, PrfA thiols may be reversibly oxidized, temporarily inactivating the protein by inhibiting DNA binding and leading to a downregulation of PrfA-regulated genes (PRG). *L. monocytogenes* could then enter the host cytosol, as PrfA activation is dispensable for vacuolar escape *in vivo*. The host cytosol is a highly reducing

environment and upon entry into this compartment, all thiols are expected to be in the reduced form. In the absence of glutathione, it is likely that coenzyme A maintains redox homeostasis in the bacterium, as it is the most abundant LMW thiol in *L. monocytogenes*. Reduced glutathione could then bind PrfA and activate transcription of PRG. This two-step activation requirement may explain why the mechanism of PrfA activation has been a mystery for over two decades; the redox changes occurring during transit through a vacuole followed by replication in the highly reducing cytosol have yet to be recapitulated *in vitro*.

Extended Data Table 1 | Strains and primers

Strain	<i>Escherichia coli</i>	Reference
NF-E1280	BL21(DE3) pET100.prfA	16
NF-E1281	BL21(DE3) pET100.prfA* (G145S)	16
DP-E6185	Rosetta pET100.prfA Quad	this work
Strain	<i>Listeria monocytogenes</i>	Reference
10403S	wt	35
DP-L3078	$\Delta actA$	36
DP-L4317	$\Delta prfA$	37
NF-L1177	PrfA* (G145S)	16
BH-3410	suicide	this work
DP-L6187	suicide <i>gshF::Tn</i>	this work
DP-L6188	$\Delta gshF$	this work
DP-L6189	$\Delta gshF + gshF$	this work
DP-L6190	PrfA* (G145S) $\Delta gshF$	this work
DP-L6191	PrfA* (G145S) <i>gshF::Tn</i>	this work
NF-L1166	PrfA* (L140F)	16
DP-L6192	PrfA* (L140F) <i>gshF::Tn</i>	this work
NF-L1214	PrfA* (Y63C)	16
DP-L6194	PrfA* (Y63C) <i>gshF::Tn</i>	this work
DP-L6195	PrfA (C/A) ₄	this work
DP-L6196	PrfA (C/A) ₄ <i>gshF::Tn</i>	this work
Primer	Oligonucleotide	Amplicon
MLR#123	CGACATAATATTTGCAGCGAC	<i>actA</i> for qPCR
MLR#124	TGCTTTCAACATTGCTATTAGG	
MLR#133	GACCCTAATCTCCGGAAGC	
MLR#134	TACAGAGTCAATCGAGTCCG	<i>gshF</i> for qPCR
MLR#121	GCGCAACAACTGAAGCAAAG	
MLR#122	CATTTGTCACTGCATCTCCG	
MLR#125	CTAACGGTCTATCTTCTAAGG	<i>hly</i> for qPCR
MLR#126	CAATAATAATTGATATAATAGCGG	
MLR#150	ACCCTTGATTTTAGTTGCCAG	
MLR#151	TGTGTAGCCCAGGTCATAAG	16S rRNA for qPCR
MLR#102	GCTTCCAAGGAGCTAAAGAGGTCCCTAGCGCC	
MLR#103	CGGGGAATTTGTATCGATAAGGAATAGATTTAAAAATTTGCTG TTATTTTG	
MLR#104	GGCCACGCGTCGACTAGTACNNNNNNNNNNCTTCT	Tn PCR Round#2
MLR#105	GGCCACGCGTCGACTAGTAC	
MLR#106	ACAATAAGGATAAATTTGAATACTAGTCTCGAGTGGGG	
KLH #1	TGAGGCATTAACATTTGTTAACGACGAT	Tn Sequencing <i>Phly</i> for DNA-binding assays
KLH #2	AACTGATTAACAAATGTTAGAGAAAAC	
		<i>PactA</i> for DNA-binding assays

References 16, 35–37 are cited in this table.

The temperature and chronology of heavy-element synthesis in low-mass stars

P. Neyskens¹, S. Van Eck¹, A. Jorissen¹, S. Goriely¹, L. Siess¹ & B. Plez²

Roughly half of the heavy elements (atomic mass greater than that of iron) are believed to be synthesized in the late evolutionary stages of stars with masses between 0.8 and 8 solar masses. Deep inside the star, nuclei (mainly iron) capture neutrons and progressively build up (through the slow-neutron-capture process^{1,2}, or s-process) heavier elements that are subsequently brought to the stellar surface by convection. Two neutron sources, activated at distinct temperatures, have been proposed: ^{13}C and ^{22}Ne , each releasing one neutron per α -particle (^4He) captured^{1–4}. To explain the measured stellar abundances^{1–7}, stellar evolution models invoking the ^{13}C neutron source (which operates at temperatures of about one hundred million kelvin) are favoured. Isotopic ratios in primitive meteorites, however, reflecting nucleosynthesis in the previous generations of stars that contributed material to the Solar System, point to higher temperatures (more than three hundred million kelvin), requiring at least a late activation of ^{22}Ne (ref. 1). Here we report a determination of the s-process temperature directly in evolved low-mass giant stars, using zirconium and niobium abundances, independently of stellar evolution models. The derived temperature supports ^{13}C as the s-process neutron source. The radioactive pair ^{93}Zr – ^{93}Nb used to estimate the s-process temperature also provides, together with the pair ^{99}Tc – ^{99}Ru , chronometric information on the time elapsed since the start of the s-process, which we determine to be one million to three million years.

We obtained high-resolution spectra of 17 S-type stars and 6 M-type stars using the high-resolution (resolving power, $\lambda/\Delta\lambda = 85,000$) HERMES spectrograph⁸ (Extended Data Table 1 and Extended Data Fig. 1). S-type stars are s-process-enriched red giants with effective temperatures in the range 3,000–4,000 K, and M-type stars are similar giant stars but show no s-process enhancement. Stellar parameters (effective temperature, surface gravity, carbon-to-oxygen ratio, s-process enhancement and metallicity $[\text{Fe}/\text{H}] = \log_{10}(N(\text{Fe})/N(\text{H}))_{\text{star}} - \log_{10}(N(\text{Fe})/N(\text{H}))_{\odot}$, where $N(\text{A})$ is the number density of element A and \odot denotes the solar value) are determined by comparing observational data with predicted spectra and photometric colours⁹ computed from a grid of dedicated model atmospheres^{10,11}. Abundances of Fe, Zr, Nb and Tc are derived, along with the corresponding errors arising from estimated uncertainties in the stellar parameters (Extended Data Tables 2–5 and Extended Data Fig. 2).

S-type stars in fact come in two varieties according to the presence or absence of Tc, which is an element with no stable isotope. ‘Extrinsic’ S stars lack Tc and are all binaries^{12,13}. The atmospheres of these giant stars contain s-process material transferred from a companion that has completed its path through the asymptotic-giant-branch phase, and the measured abundances therefore reflect the entire s-process production history. We use the $N(\text{Nb})/N(\text{Zr})$ ratio in these extrinsic S stars to derive the s-process temperature.

Mono-isotopic Nb can be produced only by β -decay of ^{93}Zr . In extrinsic S stars, the time elapsed since the end of the mass transfer from the companion is much longer than the ^{93}Zr half-life ($\tau_{1/2} = 1.53$ Myr). Consequently, the $N(\text{Zr})/N(\text{Nb})$ abundance measured today in an extrinsic S star is equal to the $N(\text{Zr})/N(^{93}\text{Zr})$ abundance at the end of the

s-process in the companion, provided that the s-process contribution dominates over the initial heavy-element abundances, which is the case in the most enriched extrinsic S stars.

As we show in the Methods (section ‘The ^{93}Zr – ^{93}Nb pair used as a thermometer’, equations (8) and (10)), the $N(\text{Zr})/N(^{93}\text{Zr})$ ratio is directly related to the neutron-capture cross-sections of the various Zr isotopes, under the approximation of local equilibrium for the s-process flow. Because these neutron-capture cross-sections depend on temperature, so does the $N(\text{Zr})/N(^{93}\text{Zr})$ ratio at the s-process site. $N(\text{Zr})/N(\text{Nb})$ is thus a measure of the s-process temperature.

Figure 1 compares the $N(\text{Zr})/N(\text{Nb})$ ratio, as predicted from the temperature-dependent neutron-capture cross-sections (squares), with those measured in the most enriched extrinsic S stars (dashed line). Figure 1 also indicates the temperature ranges for the operation of the $^{13}\text{C}(\alpha, n)^{16}\text{O}$ (compact notation for $^{13}\text{C} + ^4\text{He} \rightarrow ^{16}\text{O} + n$) and $^{22}\text{Ne}(\alpha, n)^{25}\text{Mg}$ neutron sources. To reconcile, within the 1σ uncertainty, the predicted $N(\text{Zr})/N(\text{Nb})$ ratio with that measured in extrinsic S stars, the s-process operation has to take place below $\sim 2.5 \times 10^8$ K. This is compatible with the operation of the $^{13}\text{C}(\alpha, n)^{16}\text{O}$ reaction as the neutron source but disfavors the $^{22}\text{Ne}(\alpha, n)^{25}\text{Mg}$ reaction. This confirms theoretical expectations^{1–7} that protons from the outer convective envelope diffuse in the C-rich region left after the development of the thermal pulse, which is a recurrent convective instability occurring in the He-burning shell of asymptotic-giant-branch stars. This mixing process still has a debated origin but is necessary to activate the chain of

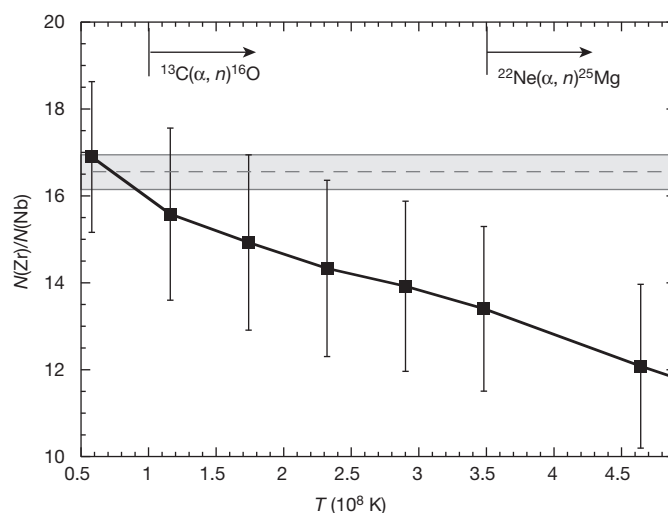


Figure 1 | The Zr–Nb pair as an s-process thermometer. We compare $N(\text{Zr})/N(\text{Nb})$, that is, $N(\text{Zr})/N(^{93}\text{Zr})$, deduced from experimental, Maxwellian-averaged cross-sections at given temperatures (squares; error bars^{14,15}, 1σ) with the value measured in the most enriched S stars (dashed line). The grey zone indicates the standard error based on orthogonal distance regression (equation (9) in the Methods). The respective temperatures above which the ^{13}C and ^{22}Ne neutron sources operate are delineated with arrows.

¹Institut d'Astronomie et d'Astrophysique, Université libre de Bruxelles (ULB), CP 226, 1050 Bruxelles, Belgium. ²Laboratoire Univers et Particules de Montpellier, Université Montpellier 2, CNRS, F-34095 Montpellier, France.

Table 1 | Derived timescales and masses for our sample of intrinsic S stars

Star name	$\log(L/L_{\odot})$	C/O	[Fe/H]	Mass loss	t_{TP} (Myr)	$t_{\text{TP,min}}$ (Myr)	$t_{\text{TP,max}}$ (Myr)	t_{S} (Myr)	M_{mod} (M_{\odot})
o ¹ Ori	3.45	0.50	−0.45	VW	1.3	0.7	2.6	0.6	2
				S	1.8	0.7	3.9	0.6	3
AA Cam	3.91	0.50	−0.40	VW	1.4	0.9	2.3	0.6	3
				S	1.6	0.7	2.7	0.4	3
KR CMa	—	0.50	−0.34	VW	1.6	0.7	2.6	0.8	3
				S	1.8	0.7	3.9	0.6	3
CSS 454	—	0.50	−0.40	VW	1.7	1.1	2.6	1.0	2
				S	1.7	1.0	3.9	1.0	2
HIP 103476	3.59	0.50	−0.01	VW	2.2	1.2	3.2	1.4	3
				S	2.5	1.5	4.0	1.3	3
AD Cyg	—	0.97	−0.05	VW	2.7	2.1	3.9	1.9	3
				S	3.0	2.2	4.0	1.8	3
NQ Pup	2.95	0.50	−0.31	VW	2.8	2.1	4.2	2.0	3
				S	3.1	2.3	4.0	1.9	3
HR Peg	3.39	0.75	0.00	VW	3.2	2.4	4.8	2.4	3
				S	3.0	2.4	3.7	1.8	3

Mass-loss rates are known to influence the stellar evolution and the minimum initial mass needed to produce intrinsic S stars. Two grids of models were thus computed, corresponding to either the VW¹⁹ or the S²⁰ mass-loss prescription. The impact on the derived ages is rather limited. t_{TP} (with associated range $t_{\text{TP,min}}-t_{\text{TP,max}}$) and t_{S} stand for the ages counted from the first thermal pulse and, respectively, from the time at which s-process elements enrich the stellar surface. M_{mod} stands for the stellar mass of the model yielding the best fit to the abundances, and M_{\odot} is the solar mass. L is the luminosity of the star²¹ and L_{\odot} is the solar luminosity. The C/O = $N(\text{C})/N(\text{O})$ and [Fe/H] values providing the best agreement with observations are also listed.

reactions $^{12}\text{C}(\text{p}, \gamma)^{13}\text{N}(\beta^+)^{13}\text{C}(\alpha, n)^{16}\text{O}$ responsible for the large neutron irradiation required to synthesize s-process elements³.

The present determination of the s-process temperature relies on a single assumption, namely that the equilibrium approximation is valid along the Zr isotopic chain, which is known to be true locally⁴. For this reason, the uncertainties in the method are mainly those in the derived abundances and in the experimental Zr cross-section^{14,15}, that is, about 5% for the stable Zr isotopes and 11% for ^{93}Zr . Reducing the ^{93}Zr error would constrain the s-process operation temperature even more.

Our result offers a long-sought confirmation of the expected—albeit still uncertain in its detailed modelling—operation of the s-process inside stars on the asymptotic giant branch. However, it contradicts previously reported higher s-process temperatures that were determined from isotopic abundance patterns in the Solar System^{1,2}, essentially by derivation from primitive carbonaceous chondrite meteorites of type CI. In contrast to the $N(\text{Zr})/N(\text{Nb})$ ratio in extrinsic S stars, which probes a single s-process site (that is, a specific star), the Solar System isotopic abundance ratios represent a mix of diverse nucleosynthetic events that took place in stars of different masses and metallicities, integrated over the Galactic evolution, until the formation of the protosolar cloud. Therefore, their interpretation is less straightforward. They lead to high s-process branching temperatures that are difficult to reconcile with nucleosynthesis in low-mass stars, where stellar models predict the s-process to occur at rather low temperatures during the interpulse phase. Instead, previous studies¹ inferring high s-process temperatures had to invoke a brief (a few years) neutron burst through $^{22}\text{Ne}(\alpha, n)^{25}\text{Mg}$ in a late hot thermal pulse, just sufficient to imprint a specific isotopic signature at the s-process branchings but not enough to affect the overall abundance distribution. The present method is insensitive to such a late and unrepresentative signature of the s-process. It instead probes the interpulse temperature, where the bulk of the s-process nucleosynthesis takes place.

Abundances in S stars act not only as a thermometer, but also as a chronometer, on the basis of the $^{93}\text{Zr}-^{93}\text{Nb}$ and $^{99}\text{Tc}-^{99}\text{Ru}$ decays in the second variety of S stars, the ‘intrinsic’ ones. These asymptotic-giant-branch stars are enriched in Tc (refs 12,16; Extended Data Figs 1 and 3). Technetium-99 is the only Tc isotope produced by the s-process, and because of its rather short half-life ($\tau_{1/2} = 0.21$ Myr), its detection is a signature of ongoing s-process nucleosynthesis. However the exploitation of Tc abundances has proved difficult^{17,18}, mainly because it is difficult to obtain reliable abundances in cool stars. The Nb and Tc abundances derived here for intrinsic S stars (Methods section ‘Abundance determinations’) now yield timescales with meaningful error bars (Table 1). Indeed, obtaining ages with sufficiently small uncertainties is always challenging when dealing with radioactive nuclides.

The derived timescales are counted from the first thermal pulse to the present time. This is shown in Fig. 2 and Extended Data Fig. 4, where observed $N(\text{Tc})/N(\text{Zr})$ and $N(\text{Nb})/N(\text{Zr})$ abundance ratios (dashed lines,

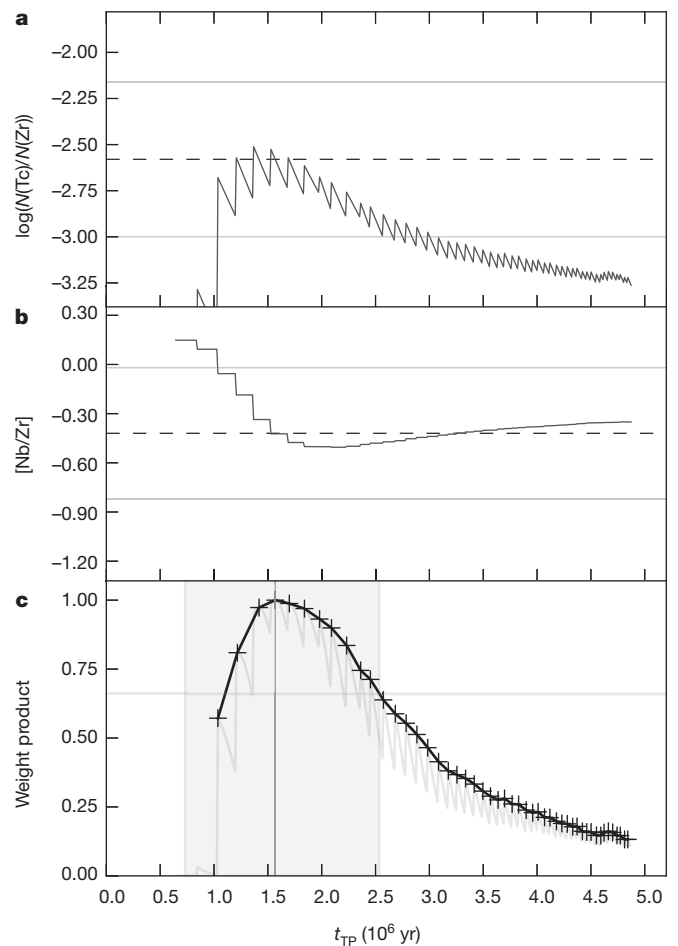


Figure 2 | KR CMa asymptotic-giant-branch timescale determination. **a, b,** We compare the predictions for $\log(N(\text{Tc})/N(\text{Zr}))$ (**a**) and $[\text{Nb}/\text{Zr}]$ (**b**) with the corresponding measured abundances (dashed lines, with statistical plus systematic errors represented as continuous lines), as functions of the time on the asymptotic giant branch (t_{TP} , counted from the first thermal pulse). **c,** The maximum weight product f_{max} (equation (5) in the Methods) gives the best-matching age (vertical grey line); the associated error bar (grey zone) is the age range covered by all models whose weight product is larger than $0.6f_{\text{max}}$ (Methods).

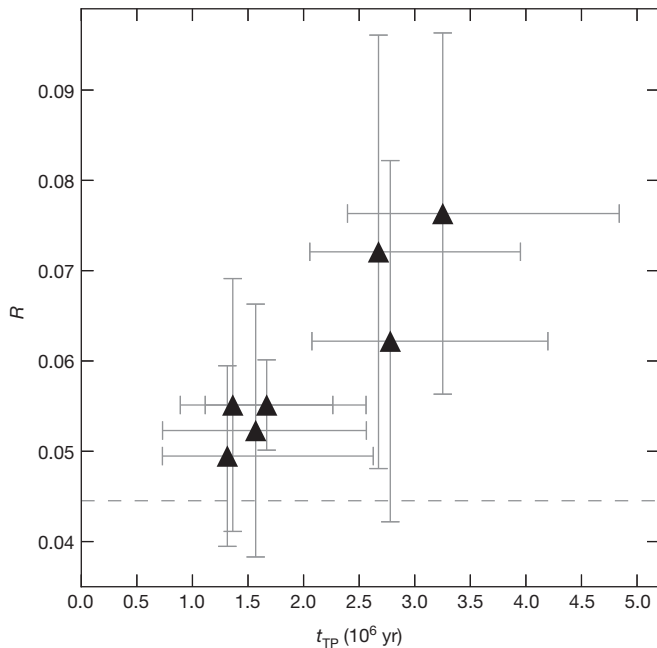


Figure 3 | Asymptotic-giant-branch timescales and infrared excess. The timescales derived from the Tc–Ru and Zr–Nb chronometers correlate well with the infrared excesses, measured by the flux ratio $R = F(12\ \mu\text{m})/F(2.2\ \mu\text{m})$. The dashed line marks the R ratio of a star without infrared excess (α Tau). Error bars on the ages are computed as explained in Fig. 2; error bars on the flux ratios are computed from the statistical error in the flux density²².

with 1σ error bars represented as solid lines) are compared with model predictions for stars on the asymptotic giant branch accounting for the production and decay of ^{99}Tc (Fig. 2a) and ^{93}Zr (Fig. 2b). In each model and for each time, a weight is assigned, separately for $N(\text{Tc})/N(\text{Zr})$ and $N(\text{Nb})/N(\text{Zr})$, depending on the difference between the observed and predicted abundance ratios (Methods section ‘Determination of TPAGB timescales’ and equation (3)). It ranges from 0 (infinite difference) to 1 (no difference). The best simultaneous agreement between measured and predicted abundances of Tc and Nb is obtained when the product of the weights is maximal; this provides the asymptotic-giant-branch evolutionary timescale (Fig. 2c) along with the corresponding model parameters (Table 1 and Extended Data Table 6).

An independent validation of the derived timescales comes from the infrared excess. As the star evolves along the asymptotic giant branch, pulsations and low surface temperatures allow dust to condense, manifesting itself as infrared excesses¹³. Therefore, we expect—and see (Fig. 3 and Extended Data Table 6)—a correlation between the asymptotic-giant-branch timescales of intrinsic S stars, counted from the first thermal pulse, and the infrared excesses $R = F(12\ \mu\text{m})/F(2.2\ \mu\text{m})$ probing the presence of circumstellar material emitting at $12\ \mu\text{m}$ with respect to the stellar photospheric flux at $2.2\ \mu\text{m}$. Although more data points are highly desirable, such a time calibration of infrared excesses is precious because infrared data are more easily obtained for these cool objects than are accurate abundances of radioactive chronometers. In future, this relation could be used as an efficient proxy for measuring asymptotic-giant-branch timescales for stars covering a limited mass range, as appears to be the case (Table 1) for intrinsic S stars.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 2 March; accepted 3 November 2014.

- Käppeler, F. The origin of the heavy elements: the s process. *Prog. Part. Nucl. Phys.* **43**, 419–483 (1999).
- Käppeler, F., Gallino, R., Bisterzo, S. & Aoki, W. The s process: nuclear physics, stellar models, and observations. *Rev. Mod. Phys.* **83**, 157–193 (2011).
- Straniero, O. *et al.* Radiative C-13 burning in asymptotic giant branch stars and s-processing. *Astrophys. J.* **440**, L85–L87 (1995).
- Goriely, S. & Mowlavi, N. Neutron-capture nucleosynthesis in AGB stars. *Astron. Astrophys.* **362**, 599–614 (2000).
- Van Eck, S., Goriely, S., Jorissen, A. & Plez, B. Discovery of three lead-rich stars. *Nature* **412**, 793–795 (2001).
- Cristallo, S. *et al.* Evolution, nucleosynthesis, and yields of low-mass asymptotic giant branch stars at different metallicities. II. The FRUITY Database. *Astrophys. J. Suppl. Ser.* **197**, 17 (2011).
- Trippella, O., Busso, M., Maiorca, E., Käppeler, F. & Palmerini, S. s-processing in AGB stars revisited. I. Does the main component constrain the neutron source in the ^{13}C pocket? *Astrophys. J.* **787**, 41 (2014).
- Raskin, G. *et al.* HERMES: a high-resolution fibre-fed spectrograph for the Mercator telescope. *Astron. Astrophys.* **526**, A69 (2011).
- Neyskens, P. *Exploring S Stars*. PhD thesis, Univ. Libre de Bruxelles (2014).
- Gustafsson, B. *et al.* A grid of MARCS model atmospheres for late-type stars. I. Methods and general properties. *Astron. Astrophys.* **486**, 951–970 (2008).
- Van Eck, S. *et al.* A grid of S stars MARCS model atmospheres. *J. Phys. Conf. Ser.* **328**, 012009 (2011).
- Van Eck, S. & Jorissen, A. The Henize sample of S stars. I. The technetium dichotomy. *Astron. Astrophys.* **345**, 127–136 (1999).
- Jorissen, A., Frayer, D. T., Johnson, H. R., Mayor, M. & Smith, V. V. S stars: infrared colors, technetium, and binarity. *Astron. Astrophys.* **271**, 463–481 (1993).
- Bao, Z. Y. *et al.* Neutron cross sections for nucleosynthesis studies. *At. Data Nucl. Data Tables* **76**, 70–154 (2000).
- Tagliente, G. *et al.* The $^{93}\text{Zr}(n,\gamma)$ reaction up to 8 keV neutron energy. *Phys. Rev. C* **87**, 014622 (2013).
- Merrill, P. W. Spectroscopic observations of stars of class S. *Astrophys. J.* **116**, 21–26 (1952).
- Smith, V. V. & Wallerstein, G. Quantitative technetium and niobium abundances in heavy-element stars. *Astrophys. J.* **273**, 742–748 (1983).
- Mathews, G. J., Takahashi, K., Ward, R. A. & Howard, W. M. Stellar technetium and niobium abundances as a measure of the lifetime of AGB stars in the third dredge-up phase. *Astrophys. J.* **302**, 410–414 (1986).
- Vassiliadis, E. & Wood, P. R. Evolution of low- and intermediate-mass stars to the end of the asymptotic giant branch with mass loss. *Astrophys. J.* **413**, 641–657 (1993).
- Schröder, K.-P. & Cuntz, M. A new version of Reimers’ law of mass loss based on a physical approach. *Astrophys. J.* **630**, L73–L76 (2005).
- Van Eck, S., Jorissen, A., Udry, S., Mayor, M. & Pernier, B. The HIPPARCOS Hertzsprung–Russell diagram of S stars: probing nucleosynthesis and dredge-up. *Astron. Astrophys.* **329**, 971–985 (1998).
- Joint IRAS Science Working Group. IRAS Catalogue of Point Sources, Version 2.0 (IPAC, 1986).

Acknowledgements P.N. acknowledges the support of a FRIA (FNRS) fellowship. S.G., L.S. and S.V.E. are FRS-FNRS Research Associates. B.P. is supported in part by the CNRS Programme National de Physique Stellaire. This work has been partly funded by an Action de recherche concertée (ARC) from the Direction générale de l’Enseignement non obligatoire et de la Recherche scientifique, Communauté française de Belgique. It was based on observations obtained using the HERMES spectrograph, which is supported by the Fund for Scientific Research of Flanders (FWO); the Research Council of K.U. Leuven; the Fonds National de la Recherche Scientifique (FRS-FNRS), Belgium; the Royal Observatory of Belgium; the Observatoire de Genève, Switzerland; and the Thüringer Landessternwarte Tautenburg, Germany.

Author Contributions S.V.E. and A.J. initiated the project. S.V.E. was supervisor of the PhD thesis of P.N., who computed the stellar atmospheric parameters, the abundances and the grid of S star model atmospheres, with help from B.P. The stellar models were computed by L.S., and the nucleosynthesis predictions were made by S.G. The text was written by S.V.E. and A.J., and edited by the other authors.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to S.V.E. (svaneck@astro.ulb.ac.be).

METHODS

Stellar parameters. Stellar parameters have been derived as follows. Narrow-band, low-resolution spectroscopic indices, probing ZrO and TiO molecular bands, are computed according to

$$B = 1 - \frac{(\lambda_{C,f} - \lambda_{C,i}) \int_{\lambda_{B,i}}^{\lambda_{B,f}} F_{\lambda} d\lambda}{(\lambda_{B,f} - \lambda_{B,i}) \int_{\lambda_{C,i}}^{\lambda_{C,f}} F_{\lambda} d\lambda} \quad (1)$$

where F_{λ} is the observed or model flux in the wavelength range $(\lambda, \lambda + d\lambda)$, with the continuum window $\lambda_{C,i}$ – $\lambda_{C,f}$ and band window $\lambda_{B,i}$ – $\lambda_{B,f}$ listed in Extended Data Table 5.

These indices, along with photometric colours (Geneva $UB_1B_2V_1V_2$ and Johnson JHK photometry), are calibrated against a grid of 3,525 dedicated MARCS model atmospheres for S stars^{10,11} covering wide ranges in effective temperature ($T_{\text{eff}} = 2,700$ – $4,000$ K), surface gravity ($\log(g \text{ cm s}^{-2}) = 0$ – 5), C/O ratio ($\text{C/O} \equiv N(\text{C})/N(\text{O}) = 0.500$ – 0.997), s-process enrichment ($[\text{s/Fe}] = 0$ – 2 , which reflects the zirconium enrichment) and metallicity ($[\text{Fe/H}] = 0$ and -0.5). Detailed synthetic spectra were calculated with state-of-the-art line lists²³ (Extended Data Table 5).

The comparison between synthetic and observed (photometric and band-strength) indices makes it possible^{9,11} to estimate T_{eff} , C/O and $[\text{s/Fe}]$ because (i) the $(V - K_0)$ and $(J - K_0)$ colours make T_{eff} and C/O separable, and (ii) the TiO and ZrO band-strength indices make T_{eff} and $[\text{s/Fe}]$ separable. Surface gravity is constrained by the $U - B_1$ and $B_2 - V_1$ colours, which probe bands from CN, CH and MgH that are sensitive to gravity. A ‘best-model-finding’ (BMF) χ^2 -minimum search between the observed and synthetic indices is then performed. A set of atmospheric parameters is thus assigned to each target star (Extended Data Table 2). Because of the strong blending of atomic lines with molecular lines, it was not possible to use standard spectroscopic methods to derive microturbulence, which has been fixed at $\chi_t = 2 \text{ km s}^{-1}$, which is a typical value for this kind of giant star.

The validity of the BMF procedure has been assessed by two consistency checks. First, for a sample of six representative S stars, there is a good agreement between the C/O ratio, $[\text{Fe/H}]$ and $[\text{s/Fe}]$ adopted for the model atmosphere and those derived from abundance determinations performed on high-resolution spectra of the same stars⁹.

Second, an accuracy test has been performed: we computed test model atmospheres representative of typical S stars, but with parameters (T_{eff} , $\log(g)$, C/O, $[\text{s/Fe}]$) falling between the corresponding parameter grid points. The resulting synthetic spectra were incorporated in the BMF procedure as real observations. Owing to the grid discretization, the model with T_{eff} , $\log(g)$, C/O and $[\text{s/Fe}]$ best matching the test model parameters is not always the first one selected by the code, but always lies within the first three selected models. Quantitatively, if χ^2_{best} represents the minimum χ^2 difference between the photometric and spectroscopic indices computed on the test model and the grid models, the model selected by the BMF procedure is always found with a χ^2 better than $1.4\chi^2_{\text{best}}$. The error in the parameters has thus been computed as the range covered by all parameters of all models with χ^2 falling in the range $[\chi^2_{\text{best}}, 1.4\chi^2_{\text{best}}]$. These parameter ranges are listed in Extended Data Table 2.

Abundance determinations. High-resolution spectra ($\lambda/\Delta\lambda = 85,000$) were obtained with the HERMES spectrograph mounted on the 1.2 m Mercator telescope at the Roque de los Muchachos Observatory (La Palma, Spain). They were reduced using the standard HERMES pipeline⁸.

Abundances (Extended Data Table 3) are derived by comparing the observed spectrum with a synthetic spectrum generated by the Turbospectrum spectral synthesis code²⁴, using the appropriate MARCS model atmosphere and atomic data as listed in Extended Data Table 5. Typical synthetic and observed spectra are illustrated in Extended Data Fig. 1.

The total uncertainty in these abundances can be estimated as

$$\begin{aligned} \sigma_{\text{tot}}^2(A) &= \sigma_A^2 + \sigma_{A,\text{model}}^2 \\ &= \sigma_A^2 + \sum_{P_i} [(\partial \log(N(A))/\partial P_i) \delta P_i]^2 \end{aligned} \quad (2)$$

where $N(A)$ is the number density of element A, σ_A is the random error (line-to-line dispersion) originating from the line properties (blends, continuum location, error on oscillator strength, and so on), $\{P_i\}$ are the model parameters, and δP_i are the typical uncertainties in the model parameters with the largest impact on the abundances: $\delta T_{\text{eff}} = 50 \text{ K}$, $\delta \log(g \text{ cm s}^{-2}) = 0.5$ and $\delta \chi_t = 0.4 \text{ km s}^{-1}$. The partial derivatives $\partial \log(N(A))/\partial P_i$ may be extracted from Extended Data Table 4, and insertion into equation (2) thus yield the following typical model errors: $\sigma_{[\text{Zr/Fe}],\text{model}} = 0.18$, $\sigma_{[\text{Nb/Fe}],\text{model}} = 0.03$, $\sigma_{[\text{Nb/Zr}],\text{model}} = 0.18$ and $\sigma_{\text{Tc/Zr},\text{model}} = 0.29$. The quadratic sum of these typical model errors and the random error specific to each star then yields the total error bar used in the abundance figures.

Extended Data Fig. 2 presents the $[\text{Zr/Fe}]$ and $[\text{Nb/Fe}]$ abundances measured in S and M stars. As expected, M giants display solar abundances, with some dispersion (± 0.15 dex in $[\text{Zr/Fe}]$ and $[\text{Nb/Fe}]$) that can be ascribed partly to uncertainties in the abundance determination (Extended Data Table 3) and partly to variations in protostellar cloud composition. S stars, however, segregate along two clearly distinct sequences (Extended Data Fig. 2), which represent the extrinsic–intrinsic dichotomy^{12,13}. On one hand, intrinsic stars have low, close-to-solar $[\text{Nb/Fe}]$ ratios, because ^{93}Zr only marginally decays during the asymptotic-giant-branch (AGB) phase. The same diagnostic applies to those carbon (C) stars (characterized by $\text{C/O} > 1$ in their atmospheres) flagged as AGB stars²⁵. On the other hand, extrinsic stars have large Nb enhancements originating from the complete decay of ^{93}Zr during the time elapsed since the end of the mass-transfer phase responsible for their pollution. This segregation in terms of Nb abundance thus represents a powerful diagnostic to separate the families of intrinsic (AGB) and extrinsic (binary) stars, in addition to the original method based on Tc detection^{12,13}.

A direct confirmation of the presence of ^{93}Zr has been attempted by searching for ^{93}Zr lines in the spectra of intrinsic S stars, but without success, probably because the ^{93}Zr s-process isotopic fraction is so small ($N(^{90}\text{Zr})/N(\text{Zr}) = 0.466$, $N(^{91}\text{Zr})/N(\text{Zr}) = 0.099$, $N(^{92}\text{Zr})/N(\text{Zr}) = 0.164$, $N(^{93}\text{Zr})/N(\text{Zr}) = 0.045$, $N(^{94}\text{Zr})/N(\text{Zr}) = 0.222$, $N(^{96}\text{Zr})/N(\text{Zr}) = 0.004$ for a $2M_{\odot}$ solar-metallicity model after ten thermal pulses) that the atomic isotopic shift does not induce any detectable line asymmetries. And contrary to a previous claim²⁶, we could not detect the ^{93}ZrO band heads at 692.5 nm, 674.2 nm, 681.15 nm, and 639.05 nm in our high-resolution spectra.

Nucleosynthesis computations. As it ascends the AGB, a star is subject to recurrent convective instabilities that develop in the He-burning shell, the thermal pulses. The corresponding AGB part is called the TPAGB. Following these He-shell flashes, the convective envelope is able to penetrate the region previously occupied by the thermal pulse, bringing to the surface the products of a rich nucleosynthesis. This process is called the third dredge-up. Nucleosynthesis by the s-process occurs during the TPAGB phase, when protons from the envelope diffuse downwards into the C-rich region of the pulse. The subsequent chain of reactions $^{12}\text{C}(p, \gamma)^{13}\text{N}(\beta^+)^{13}\text{C}(\alpha, n)^{16}\text{O}$ leads to the formation of a ^{13}C pocket and to the release of neutrons^{1–7}.

At solar metallicity, the minimum initial mass for the appearance of third dredge-ups is $M \approx 1.5M_{\odot}$ – $1.7M_{\odot}$, depending on the assumed mass-loss prescription. In this paper, we computed with the STAREVOL code²⁷ a grid of models covering stellar masses $M_{\text{mod}} = 1.7M_{\odot}$, $2.0M_{\odot}$, $3.0M_{\odot}$ and $4.0M_{\odot}$, metallicities $Z = 0.008$ ($[\text{Fe/H}] = -0.5$) and $Z = 0.012$ ($[\text{Fe/H}] = 0.0$, that is, solar abundances²⁸), and initial $[\text{Nb/Fe}]$ values of -0.15 , 0 and 0.15 dex. The Vassiliadis–Wood¹⁹ (VW) and Schröder²⁰ (S) mass-loss prescriptions were both implemented. With increasing mass, the temperature at the base of the convective envelope rises and hot-bottom burning is activated. Under such conditions, corresponding to stars with $M > \sim 4M_{\odot}$ at solar metallicity, the efficiency of the s-process nucleosynthesis is strongly reduced because the diffusing protons are burnt through $^{12}\text{C}(p, \gamma)^{13}\text{N}(\beta^+)^{13}\text{C}(p, \gamma)^{14}\text{N}$ while travelling into the C-rich region and the ^{13}C pocket is contaminated with some ^{14}N . The result is a dramatic reduction in the production of heavy elements by the s-process²⁹. Additionally, stars with $M > \sim 4M_{\odot}$ are able to activate the $^{22}\text{Ne}(\alpha, n)^{25}\text{Mg}$ neutron source at the base of the thermal pulse, which results in an s-process signature quite different from that of lower-mass stars³⁰, and which, as we checked on our sample, does not agree with the measured distribution of abundances in S stars.

The s-process nucleosynthesis is obtained by post-processing^{4,5,29} the stellar models with an extended nuclear network using the up-to-date cross-sections available from the NETGEN nuclear library (<http://www.astro.ulb.ac.be/Netgen/form.html>). Large uncertainties exist in the formation of the ^{13}C pocket, because the efficiency of the proton mixing in the ^{12}C layer, which is responsible for the formation of the ^{13}C pocket, cannot yet be derived from first principles. Therefore, the extent of the partial mixing region is described by a free parameter λ_{pm} , characterizing the ratio between the mass of the proton injection region and that of the thermal pulse. In our computations, the values $\lambda_{\text{pm}} = 2\%$, 3% , 5% , 10% and 30% have been used. These values are larger than what would be expected from a standard overshooting prescription^{4,29}, but are required to match the level of observed s-process enrichments.

The measured abundances and those predicted by our model (Extended Data Fig. 2, shaded areas) are fully consistent with each other, with some sensitivity to the s-process modelling, mostly via the mixing parameter λ_{pm} .

Determination of TPAGB timescales. The decay of ^{93}Zr and ^{99}Tc is used to determine stellar timescales, counted from the beginning of the TPAGB (that is, from the first thermal pulse), consistently taking into account the temperature dependence of the β -decay rates of ^{99}Tc and ^{93}Zr in our nucleosynthesis computations³¹.

It is necessary to quantify the agreement between the predicted and measured abundances of the two chronometers Tc and Nb. At time t , the predicted Tc abundance for a star of initial mass M_{mod} , metallicity $[\text{Fe/H}]$, partial mixing parameter λ_{pm} and initial Nb abundance $N_0(\text{Nb})$, is denoted $N_i(\text{Tc}; t)$, where i represents the set of model parameters $\{M_{\text{mod}}, \lambda_{\text{pm}}, N_0(\text{Nb}), [\text{Fe/H}]\}$. For each model i , each time

t is assigned a weight W_i depending on the difference between $N_i(\text{Tc}; t)$ and the observed Tc abundance. More precisely, this weight follows a Gaussian distribution that is a function of the abundance x (as illustrated on the right of Extended Data Fig. 4a, b):

$$W_{\text{Tc}}(x) = e^{-\frac{1}{2} \left(\frac{x - \mu(\text{Tc})}{\sigma(\text{Tc})} \right)^2} \quad (3)$$

Here $\mu(\text{Tc})$ and $\sigma(\text{Tc})$ are the measured Tc abundance and its associated uncertainty, respectively. A weight function $f_{\text{Tc},i}(t)$ is computed such that, for each $x = N_i(\text{Tc}; t)$

$$f_{\text{Tc},i}(t) = W_{\text{Tc}}(x) \quad (4)$$

A similar function $f_{\text{Nb},i}(t)$ is computed along the same lines as for $f_{\text{Tc},i}(t)$, and the two are plotted in Extended Data Fig. 4c. The weight product function, for the Tc and Zr chronometers, is

$$f_i(t) = f_{\text{Tc},i}(t) f_{\text{Nb},i}(t) \quad (5)$$

The grey curves in Extended Data Fig. 4d show this weight product $f_i(t)$ for various models. The weight product corresponding to the best-fitting model is plotted in green. Its maximum corresponds to the stellar age on the TPAGB ($f_{\text{max}} = f_{\text{best}}(t_{\text{best}})$, vertical line in Extended Data Fig. 4d). The uncertainty in the TPAGB age ($t_{\text{TP,max}}$ and $t_{\text{TP,min}}$ in Table 1 and Extended Data Table 6) is computed as the time interval $[t_{\text{min}}, t_{\text{max}}]$ such that, for all $t \in [t_{\text{min}}, t_{\text{max}}]$, there exists i such that $f_i(t) > \alpha f_{\text{best}}(t_{\text{best}})$, where α has been set to 0.6 by analogy with the 1σ deviation from the maximum of a Gaussian (horizontal line in Extended Data Fig. 4d). The typical uncertainty in the stellar masses (M_{mod} in Table 1 and Extended Data Table 6) is derived in a similar way, by considering all models with a weight product function peaking at values larger than $\alpha f_{\text{best}}(t_{\text{best}})$.

As an internal consistency check, we verified that the timescales derived from Tc and Nb overabundances were always consistent with the timescales derived from the minimum and maximum numbers of thermal pulses required to provide the measured overabundances of other s-process elements (Y, Zr, La and Ce). Furthermore, good agreement is always found between the ^{99}Tc and ^{93}Zr chronometers, thus providing well-defined TPAGB timescales (Table 1 and Extended Data Table 6), except for the star NQ Pup, which is abnormal in other respects as well. In particular, as illustrated by its location in the Hertzsprung–Russell diagram (Extended Data Fig. 3), it is the intrinsic star with the lowest luminosity²¹ ($\log(L/L_{\odot}) = 2.96$). Its intrinsic nature is, however, secured from its Tc and Nb content. If NQ Pup were removed from our stellar sample, it would not alter the conclusions of the present study: TPAGB timescale correlation with infrared excess (and with C/O ratio; see Table 1), would even be tighter, and the s-process operation temperature would remain unchanged.

We note that AA Cam, with a TPAGB age of 1.4×10^6 yr making it one of the least evolved TPAGB stars in our sample, has the smallest mass-loss rate, as derived from CO data³² in a sample of ten intrinsic S stars.

Finally, we note that HR Peg is the only star of our sample displaying a strong Li line, and is presumably Li-rich. This is consistent with the fact that it is an intrinsic S star; indeed, Li has been shown to be a useful diagnostic in the intrinsic–extrinsic classification³². It is believed to be produced in higher-mass stars, but our $4M_{\odot}$ models are never found to be compatible with the measured s-process abundances, in contrast to $M \leq 3M_{\odot}$ models.

The ^{93}Zr – ^{93}Nb pair used as a thermometer. We first define the s-process dilution factor as

$$f = \frac{\Delta M}{M + \Delta M} \quad (6)$$

where M is the amount of mass in which ΔM is diluted. If applied to the atmosphere of an intrinsic star, ΔM is the mass of matter freshly processed by the s-process and dredged up in the stellar envelope of mass M . If applied to the atmosphere of an extrinsic star, the dilution factor combines the dilution of s-processed matter in the primary (intrinsic) star envelope, the mass-accretion efficiency and the dilution of this accreted mass in the secondary (extrinsic) star envelope.

If $X_0(\text{Zr})$, $X_s(\text{Zr})$ and $X(\text{Zr})$ are respectively the initial Zr mass fraction in mass M , the Zr mass fraction in the s-processed matter and the final Zr mass fraction in mass $M + \Delta M$, we have

$$X(\text{Zr}) = (1-f)X_0(\text{Zr}) + fX_s(\text{Zr})$$

and, in terms of number densities ($N = X/A$, where A is the mass number)

$$N(\text{Zr}) = (1-f)N_0(\text{Zr}) + fN_s(\text{Zr}) \quad (7)$$

With the usual bracket notation

$$\left[\frac{\text{Zr}}{\text{Fe}} \right] = \log \left(\left((1-f) \frac{N_0(\text{Zr})}{N_{\odot}(\text{Zr})} + f \frac{N_s(\text{Zr})}{N_{\odot}(\text{Zr})} \right) \frac{N_{\odot}(\text{Fe})}{N_s(\text{Fe})} \right)$$

A similar expression holds for Nb; therefore:

$$\left[\frac{\text{Zr}}{\text{Fe}} \right] = \left[\frac{\text{Nb}}{\text{Fe}} \right] + \log \left(\frac{(1-f) \frac{N_0(\text{Zr})}{N_{\odot}(\text{Zr})} + f \frac{N_s(\text{Zr})}{N_{\odot}(\text{Zr})}}{(1-f) \frac{N_0(\text{Nb})}{N_{\odot}(\text{Nb})} + f \frac{N_s(\text{Nb})}{N_{\odot}(\text{Nb})}} \right)$$

Asymptotically, when the abundance of the s-processed material dominates over the initial composition ($fN_s \gg (1-f)N_0$ in equation (7))

$$\left[\frac{\text{Zr}}{\text{Fe}} \right] = \left[\frac{\text{Nb}}{\text{Fe}} \right] + \log \frac{N_s(\text{Zr})}{N_s(\text{Nb})} - \log \frac{N_{\odot}(\text{Zr})}{N_{\odot}(\text{Nb})}$$

We define ω^* as the ratio of the Zr abundance to the ^{93}Zr abundance in the s-processed material:

$$\omega^* = \frac{N_s(\text{Zr})}{N_s(^{93}\text{Zr})} \quad (8)$$

When all ^{93}Zr has decayed into mono-isotopic Nb, we have

$$\omega^* = \frac{N_s(\text{Zr})}{N_s(\text{Nb})}$$

and

$$\left[\frac{\text{Zr}}{\text{Fe}} \right] = \left[\frac{\text{Nb}}{\text{Fe}} \right] + \log(\omega^*) - \log \frac{N_{\odot}(\text{Zr})}{N_{\odot}(\text{Nb})} \quad (9)$$

The value $\log(\omega^*) - \log(N_{\odot}(\text{Zr})/N_{\odot}(\text{Nb}))$ can thus be estimated using equation (9) and Extended Data Fig. 2, by fitting a straight line of slope 1 through the points for which the approximation $fN_s \gg (1-f)N_0$ is valid. This occurs in extrinsic S stars where all ^{93}Zr has decayed into mono-isotopic Nb and where strong s-process pollution has erased any pristine abundance profile. Therefore, the two extrinsic stars richest in Zr provide a measurement of ω^* because for these stars $N(\text{Zr})/N(\text{Nb}) \approx N_s(\text{Zr})/N_s(\text{Nb}) \equiv \omega^*$.

Furthermore, nuclear physics provides the temperature dependence of ω^* : to a very good approximation, the s-process flow is in local equilibrium within a given isotopic chain. In the case of Zr, this is true at least between ^{90}Zr and ^{94}Zr where no branching point occurs, such that $N_s(^A\text{Zr})\langle\sigma_A\rangle$ is constant within the Zr chain ($\langle\sigma_A\rangle$ is the Maxwellian-averaged neutron-capture cross-section^{14,15} of ^AZr). Thus

$$\omega^* = \langle\sigma_{93}\rangle \left[\frac{1}{\langle\sigma_{90}\rangle} + \frac{1}{\langle\sigma_{91}\rangle} + \frac{1}{\langle\sigma_{92}\rangle} + \frac{1}{\langle\sigma_{94}\rangle} \right] \quad (10)$$

Because ω^* is temperature dependent through the temperature dependence of the individual cross-sections, it can be used to constrain the temperature at which the Zr production took place.

We find a temperature for the s-process less than $\sim 2.5 \times 10^8$ K, favouring the $^{13}\text{C}(\alpha, n)^{16}\text{O}$ neutron source over the $^{22}\text{Ne}(\alpha, n)^{25}\text{Mg}$ one. The deduced temperature range (Fig. 1) is an upper estimate: if we had considered more stars as being dominated by the s-process pollution, the y intercept of equation (9) (that is, $\log(\omega^*) - \log(N_{\odot}(\text{Zr})/N_{\odot}(\text{Nb}))$) would have been larger, and the derived temperature even lower (Fig. 1), strengthening the case for the $^{13}\text{C}(\alpha, n)^{16}\text{O}$ neutron source. Finally, the derived temperature is independent of the solar abundances and their associated errors, as can be seen from equation (9), because the square-bracket notation corresponds to abundance ratios normalized to their solar values, and these cancel out.

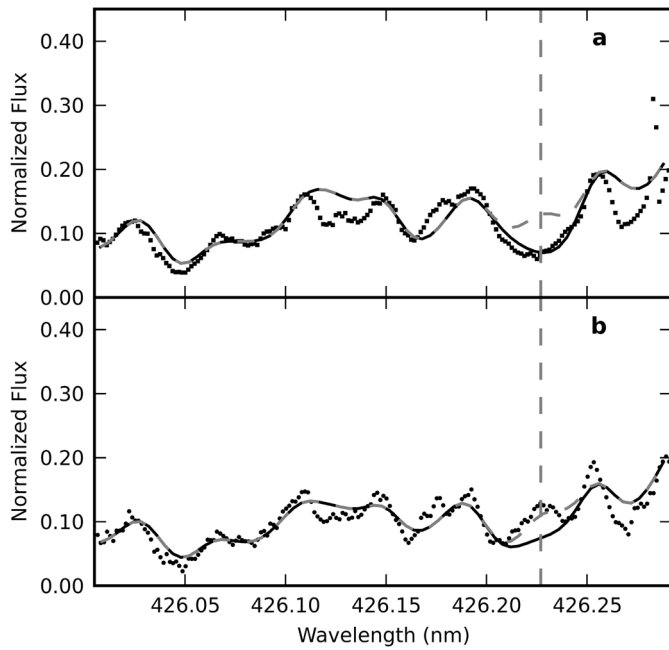
Taking mass loss into account in the dilution factor (equation (6)) will increase f , and so the condition $fN_s \gg (1-f)N_0$ will be satisfied even sooner. Because only this asymptotic regime is used to constrain the s-process operation temperature, our temperature determination is unaffected by a diminution of the envelope mass.

Previous estimates of the s-process temperature were based on the analysis of Solar System abundances^{1,28}. These were derived essentially from primitive carbonaceous chondrite meteorites of type CI, where most s-process branchings lead to a high temperature of the order of 3×10^8 K for the s-process operation, thus pointing towards the activation of the $^{22}\text{Ne}(\alpha, n)^{25}\text{Mg}$ neutron source. The higher temperatures so inferred were ascribed to a short, final neutron burst from the $^{22}\text{Ne}(\alpha, n)^{25}\text{Mg}$ neutron source operating in a late, hot thermal pulse. However, these meteoritic abundances are notoriously difficult to interpret^{1–3}, for the following reasons. First, they are a mix of diverse nucleosynthetic events taking place in stars of different masses and metallicities. Second, they are potentially sensitive to

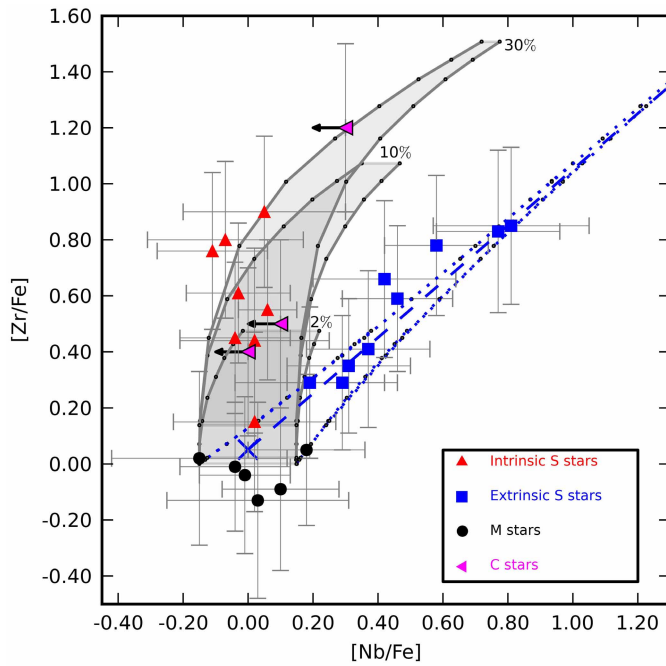
differential depletion effects during grain condensation and contamination with Solar System material³⁴. Third, the temperatures are deduced from branchings (that is, unstable nuclei for which there is competition between neutron capture and β -decay) along the s-process path; however, some branchings are sensitive not only to temperature, but also to poorly known neutron and electron densities. Finally, the branching neutron-capture cross-sections and the temperature dependence of the β -decays are not always well constrained experimentally.

Our work shows that the event associated with $^{22}\text{Ne}(\alpha, n)^{25}\text{Mg}$, despite imprinting its signature in the material incorporated in the Solar System carbonaceous chondrites, is not representative of the s-process in low-mass stars, which is instead powered by the $^{13}\text{C}(\alpha, n)^{16}\text{O}$ neutron source.

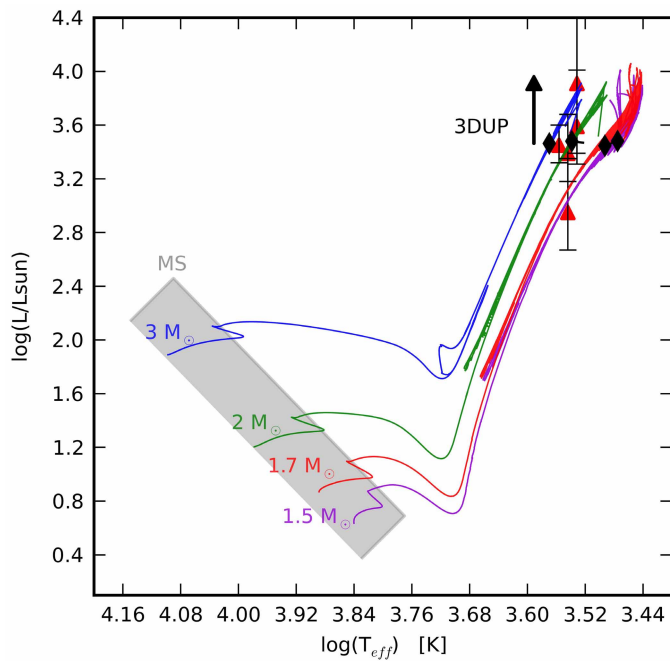
23. Kupka, F., Piskunov, N., Ryabchikova, T. A., Stempels, H. C. & Weiss, W. W. VALD-2: Progress of the Vienna atomic line data base. *Astron. Astrophys.* **138**, 119–133 (1999).
24. Alvarez, R. & Plez, B. Near-infrared narrow-band photometry of M-giant and Mira stars: models meet observations. *Astron. Astrophys.* **330**, 1109–1119 (1998).
25. de Laverny, P. *et al.* Chemical analysis of carbon stars in the Local Group. *Astron. Astrophys.* **446**, 1107–1118 (2006).
26. Lambert, D. L., Smith, V. V., Busso, M., Gallino, R. & Straniero, O. The chemical composition of red giants. IV. The neutron density at the s-process site. *Astrophys. J.* **450**, 302–317 (1995).
27. Siess, L. Evolution of massive AGB stars. I. Carbon burning phase. *Astron. Astrophys.* **448**, 717–729 (2006).
28. Asplund, M., Grevesse, N. & Sauval, A. J. in *Cosmic Abundances as Records of Stellar Evolution and Nucleosynthesis* (eds Barnes, T. G. III & Bash, F. N.) 25–38 (Astronomical Society of the Pacific, 2005).
29. Goriely, S. & Siess, L. S-process in hot AGB stars: a complex interplay between diffusive mixing and nuclear burning. *Astron. Astrophys.* **421**, L25–L28 (2004).
30. García-Hernández, D. A. *et al.* Rubidium-rich asymptotic giant branch stars. *Science* **314**, 1751–1754 (2006).
31. Takahashi, K. & Yokoi, K. Beta-decay rates of highly ionized heavy atoms in stellar interiors. *At. Data Nucl. Data Tables* **36**, 375–409 (1987).
32. Wallerstein, G., Balick, B., Alcolea, J., Bujarrabal, V. & Vanture, A. D. Carbon isotopic abundance ratios in S-type stars. *Astron. Astrophys.* **535**, A101 (2011).
33. Savina, M. R. *et al.* Extinct technetium in silicon carbide stardust grains: implications for stellar nucleosynthesis. *Science* **303**, 649–652 (2004).
34. Kashiv, Y. *et al.* Extinct ^{93}Zr in single presolar SiC Grains from low mass asymptotic giant branch stars and condensation from Zr-depleted gas. *Astrophys. J.* **713**, 212–219 (2010).
35. Stephenson, C. B. *A General Catalogue of Galactic S Stars* 2nd edn, Publications of the Warner & Swasey Observatory 3 (1984).
36. Biémont, E., Grevesse, N., Hannaford, P. & Lowe, R. M. Oscillator strengths for Zr I and Zr II and a new determination of the solar abundance of zirconium. *Astrophys. J.* **248**, 867–873 (1981).



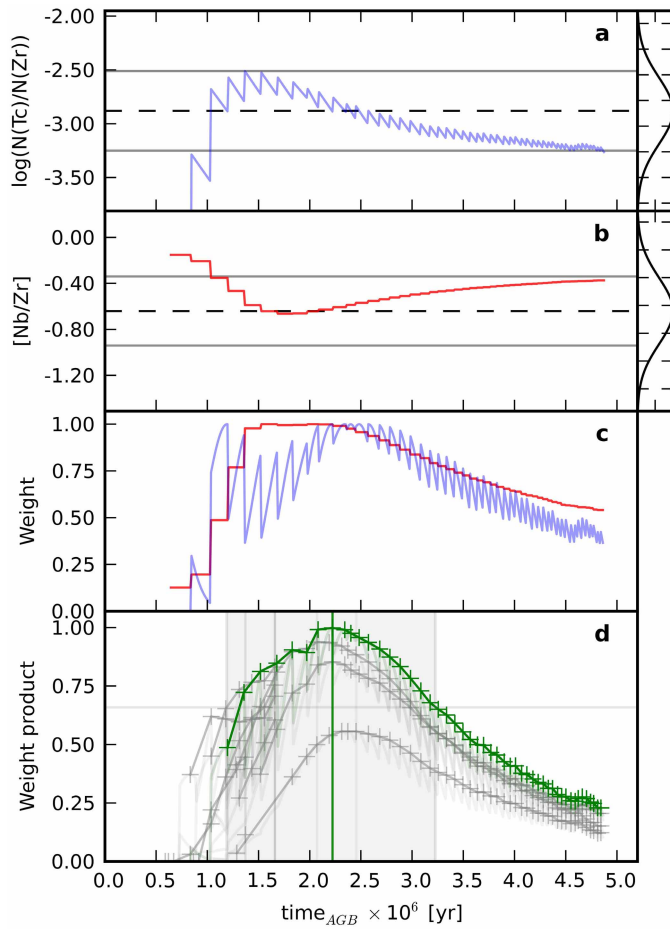
Extended Data Figure 1 | Observed and synthetic spectra around the Tc I 426.227 nm line. **a**, Observed spectrum of the intrinsic, Tc-rich star NQ Pup (dots) along with spectral syntheses with $\log(\epsilon_{\text{Tc}}) = \log(N(\text{Tc})/N(\text{H})) + 12.0 = -0.20$ (continuous line; see Extended Data Table 3) and no Tc (dashed line). **b**, Same as **a**, but for the extrinsic, Tc-poor star V613 Mon; spectral syntheses correspond to $\log(\epsilon_{\text{Tc}}) = -0.30$ (continuous line) and no Tc (dashed line).



Extended Data Figure 2 | Predicted versus measured abundances. We show that Nb and Zr abundances for intrinsic S stars and carbon stars (with only upper limits on the Nb abundance²⁵) are consistent with nucleosynthesis predictions (shaded areas) for a $2.0M_{\odot}$ model with a solar initial Nb abundance (± 0.15 dex) and different λ_{pm} values²⁹ (as labelled). Once all ^{93}Zr has decayed to ^{93}Nb , the models (dotted lines) reproduce the extrinsic S stars. The error bars represent the line-to-line dispersion and the systematic uncertainty from the models. The dashed line is a fit of equation (9) through the two most enriched extrinsic stars.



Extended Data Figure 3 | Location of intrinsic S stars in the Hertzsprung-Russell diagram. We show the location of intrinsic (that is, Tc-rich) S stars (triangles) from our sample in the stellar luminosity/effective temperature diagram²¹, along with low-mass evolutionary tracks (starting at the zero-age main sequence). The main sequence (MS) is indicated. The black diamond on each evolutionary track indicates where the third dredge-up event starts.



Extended Data Figure 4 | Timescale determination for HIP 103476.

a, b, $N(\text{Tc})/N(\text{Zr})$ (**a**) and $[\text{Nb}/\text{Zr}]$ (**b**) predictions from the best-matching model (as judged from panel **d**, with the same meaning as in Fig. 2). For reasons of clarity we plot only these predictions. **c,** The deviation between predicted (solid line) and observed (dashed line) abundances in **a** and **b**, converted into weight functions $f_{\text{Tc},i}(t)$ (blue) and $f_{\text{Nb},i}(t)$ (red) (equation (4)). **d,** The maximum value reached by their product (green) (equation (5)) provides the best age estimate (vertical line). Weight product functions corresponding to less satisfactory models are plotted as grey lines.

Extended Data Table 1 | Summary of observations

Star name	Date Obs.	T (s)	S/N
CSS 454	2011, Feb 17	2800	40
α^1 Ori	2011, Apr 10	60	74
KR CMa	2011, Apr 6	1200	50
NQ Pup	2010, Feb 20	800	100
AD Cyg	2009, May 19	2000	78
HR Peg	2009, Jul 18	420	44
AA Cam	2010, Apr 3	600	39
HIP 103476	2009, Jun 20	1500	130
BD Cam	2009, Sep 17	120	109
HD 119667	2011, Apr 10	995	57
V613 Mon	2009, Sep 24	800	50
HD 191226	2009, May 19	500	78
HD 191589	2011, Apr 7	533	55
V530 Lyr	2011, Apr 2	647	39
CPD -19°1672	2011, Jan 11	2600	50
HR 363	2009, Jul 27	200	84
V1261 Ori	2009, Nov 29	500	55
RR UMi	2010, Feb 7	300	78
V 762 Cas	2011, Aug 11	400	110
V 465 Cas	2010, Aug 23	400	100
μ Gem	2012, Mar 9	900	212
ρ Per	2012, Jul 8	60	141
RZ Ari	2013, Feb 9	100	39

The wavelength range of the HERMES spectrograph^a is 380–900 nm. T is the exposure time; the signal-to-noise ratio (S/N) is estimated at around 500 nm.

Extended Data Table 2 | Adopted atmospheric parameters for the S and M stars of the present study

Star	T_{eff} (K)			$\log(g/\text{cm s}^{-2})$			[Fe/H]			C/O			[s/Fe]			χ^2_{best}
	min	max		min	max		min	max		min	max		min	max		
CSS 454	3500	3500	3500	0.	0.	0.	-0.5	-0.5	-0.5	0.50	0.50	0.50	1.	1.	1.	0.0278
σ^1 Ori	3600	3500	3600	0.	0.	1.	0.0	-0.5	0.0	0.50	0.50	0.92	0.	0.	1.	0.0911
KR CMa	3400	3300	3400	0.	0.	1.	0.0	-0.5	0.0	0.50	0.50	0.75	0.	0.	0.	0.0392
NQ Pup	3500	3500	3500	1.	0.	1.	-0.5	-0.5	-0.5	0.50	0.50	0.75	1.	1.	1.	0.0193
AD Cyg	3300	3100	3300	0.	0.	1.	0.0	-0.5	0.0	0.97	0.97	0.99	1.	1.	1.	0.1450
HR Peg	3500	3500	3500	1.	1.	2.	-0.5	-0.5	-0.5	0.75	0.50	0.75	1.	1.	1.	0.0187
AA Cam	3400	3300	3400	1.	1.	2.	0.0	-0.5	0.0	0.50	0.50	0.75	0.	0.	0.	0.0233
HIP 103476	3400	3400	3500	1.	0.	1.	-0.5	-0.5	-0.5	0.50	0.50	0.75	1.	1.	1.	0.0628
BD Cam	3600	3600	3600	1.	0.	1.	0.0	0.0	0.0	0.50	0.50	0.50	1.	1.	1.	0.0216
HD 119667	3700	3700	3700	1.	1.	2.	-0.5	-0.5	-0.5	0.50	0.50	0.50	1.	1.	1.	0.0200
V613 Mon	3500	3500	3500	0.	0.	1.	-0.5	-0.5	-0.5	0.75	0.75	0.95	1.	1.	1.	0.0987
HD 191226	3600	3500	3700	1.	0.	1.	-0.5	-0.5	0.0	0.75	0.50	0.99	0.	0.	1.	0.2941
HD 191589	3700	3600	3800	1.	0.	1.	0.0	-0.5	0.0	0.50	0.50	0.97	0.	0.	0.	0.0462
V530 Lyr	3500	3500	3600	0.	0.	0.	-0.5	-0.5	0.0	0.50	0.50	0.75	0.	0.	0.	0.0510
CPD -19°1672	3600	3600	3700	1.	0.	1.	-0.5	-0.5	0.0	0.75	0.50	0.95	0.	0.	0.	0.0404
HR 363	3600	3600	3600	1.	0.	1.	-0.5	-0.5	-0.5	0.50	0.50	0.50	1.	1.	1.	0.0351
V1261 Ori	3600	3600	3700	1.	1.	2.	-0.5	-0.5	0.0	0.50	0.50	0.75	1.	1.	1.	0.0558
RR UMi	3500	3400	3500	1.	1.	3.	0.0	0.0	0.0	0.50	0.50	0.50	0.	0.	0.	0.0154
V762 Cas	3800	3700	3800	0.	0.	0.	0.0	-0.5	0.0	0.50	0.50	0.95	0.	0.	0.	0.0199
V465 Cas	3400	3300	3400	1.	2.	2.	-0.5	-0.5	-0.5	0.50	0.50	0.50	0.	0.	0.	0.0146
μ Gem	3600	3600	3600	1.	0.	1.	0.0	0.0	0.0	0.50	0.50	0.75	0.	0.	0.	0.0807
ρ Per	3500	3500	3500	1.	0.	1.	-0.5	-0.5	-0.5	0.50	0.50	0.50	0.	0.	0.	0.0365
RZ Ari	3400	3300	3500	0.	0.	2.	-0.5	-0.5	0.0	0.50	0.50	0.75	0.	0.	0.	0.1212

The 'min' and 'max' columns list the extremes of the parameter ranges covered by the various best-fitting model atmospheres (see text). The last column lists χ^2_{best} , that is, the minimum χ^2 difference between spectral and photometric indices computed from the observed spectrum and from a synthetic spectrum corresponding to the listed parameters.

Extended Data Table 3 | Derived abundances for S, M and C stars

Star	Spec.	Typ.	e/i	[Fe/H]	$\sigma_{[\text{Fe}/\text{H}]}$	[Nb/Fe]	$\sigma_{[\text{Nb}/\text{Fe}]}$	[Zr/Fe]	$\sigma_{[\text{Zr}/\text{Fe}]}$	$\log \epsilon_{\text{Tc}}$	$\sigma(\log \epsilon_{\text{Tc}})$
CSS 454	S		i	-0.40	0.12	-0.07	0.24	0.80	0.21	0.20	0.20
<i>o</i> ¹ Ori	M3S		i	-0.45	0.07	-0.04	0.17	0.45	0.20	-0.05	0.20
KR CMa	M4S		i	-0.34	0.11	0.02	0.23	0.44	0.27	0.10	0.20
NQ Pup	S5/2		i	-0.31	0.13	-0.11	0.16	0.76	0.21	-0.20	0.20
AD Cyg	S5/5		i	-0.05	0.14	0.05	0.24	0.90	0.20	0.30	0.20
HR Peg	S4/1		i	0.00	0.06	0.06	0.08	0.55	0.18	-0.10	0.10
AA Cam	M5S		i	-0.04	0.05	0.02	0.24	0.15	0.25	0.10	0.20
HIP 103476	MS		i	-0.01	0.06	-0.03	0.16	0.61	0.18	0.30	0.20
BD Cam	S4/2		e	-0.03	0.06	0.58	0.16	0.78	0.18	-	-
HD 119667	M1wkS		e	0.01	0.07	0.46	0.17	0.59	0.18	-	-
V613 Mon	S3/2		e	-0.26	0.15	0.42	0.22	0.66	0.21	-	-
HD 191226	M1-3S		e	-0.28	0.10	0.31	0.19	0.35	0.16	-	-
HD 191589	S		e	0.01	0.06	0.29	0.17	0.29	0.15	-	-
V530 Lyr	S3/1		e	-0.16	0.09	0.19	0.23	0.29	0.20	-	-
CPD -19°1672	S3		e	-0.01	0.06	0.37	0.19	0.41	0.21	-	-
HR 363	S3/2		e	-0.38	0.12	0.77	0.19	0.83	0.22	-	-
V1261 Ori	S4,1		e	-0.22	0.15	0.81	0.23	0.85	0.22	-	-
RR UMi	M5III			-0.06	0.06	-0.01	0.13	-0.04	0.22	-	-
V762 Cas	M3			-0.06	0.09	-0.04	0.17	-0.01	0.14	-	-
V465 Cas	M3			0.03	0.05	0.03	0.27	-0.13	0.21	-	-
μ Gem	M3III			-0.15	0.06	0.18	0.18	0.05	0.20	-	-
ρ Per	M4II			-0.06	0.09	0.10	0.17	-0.09	0.23	-	-
RZ Ari	M6III			-0.22	0.20	-0.15	0.27	0.02	0.25	-	-
[IGI95] C1	C			-0.8	0.3	≤ 0.10	-	0.50	0.30	-	-
[IGI95] C3	C			-0.5	0.3	≤ 0.30	-	1.20	0.30	-	-
BMB-B 30	C			-1.0	0.3	≤ 0.00	-	0.40	0.30	-	-

The S star spectral types are from ref. 35 and the Cstar data are from ref. 25. The usual spectroscopic notation is adopted: for elements A and B, $\log(\epsilon_A) = \log(N(A)/N(H)) + 12.0$ and $[A/B] = \log(N(A)/N(B)) - \log(N(A)/N(B))_{\odot}$. The e/i column indicates whether the star abundance peculiarities are of extrinsic (e) or intrinsic (i) origin.

Extended Data Table 4 | Sensitivity of Fe, Nb, Zr and Tc abundances to model atmosphere parameter variations (effective temperature, gravity and microturbulence χ_t)

Element	$\Delta T_{\text{eff}} = +100 \text{ K}$	$\Delta \log(g/\text{cm s}^{-2}) = +0.5$	$\Delta \chi_t = -0.4 (\text{km s}^{-1})$
$\Delta \log N(\text{Fe})$	-0.10	+0.14	+0.07
$\Delta \log N(\text{Nb})$	-0.15	+0.15	+0.07
$\Delta \log N(\text{Zr})$	-0.13	+0.04	-0.08
$\Delta \log N(\text{Tc})$	+0.10	+0.30	+0.00
$\Delta [\text{Zr}/\text{Fe}]$	-0.03	-0.10	-0.15
$\Delta [\text{Nb}/\text{Fe}]$	-0.05	+0.01	+0.00
$\Delta [\text{Nb}/\text{Zr}]$	-0.02	+0.11	+0.15
$\Delta \log(N(\text{Tc})/N(\text{Zr}))$	+0.23	+0.26	+0.08

The nominal model atmosphere has $T_{\text{eff}} = 3,500 \text{ K}$, $\log[g (\text{cm s}^{-2})] = 1$, $\chi_t = 2 \text{ km s}^{-1}$, $[\text{Fe}/\text{H}] = -0.5 \text{ dex}$, $\text{C}/\text{O} = 0.752$ and $[\text{s}/\text{Fe}] = 1.00 \text{ dex}$.

Extended Data Table 5 | Molecular and atomic lines used for computation of the stellar parameters, and atomic lines used for the abundance determinations

a					b							
Band	$\lambda_{B,i}$	$\lambda_{B,f}$	$\lambda_{C,i}$	$\lambda_{C,f}$	Element	λ (nm)	$\chi_{exc}(eV)$	$\log(gf)$	Element	λ (nm)	$\chi_{exc}(eV)$	$\log(gf)$
ZrO	574.8	575.7	580.0	581.0	Fe I	872.914	3.41	-2.87	Tc I	426.2297	0.00	-1.297
ZrO	634.2	637.3	646.2	647.2		871.039	4.91	-0.54		426.2320	0.00	-1.297
ZrO	637.8	638.2	646.2	647.2		869.870	2.99	-3.45		426.2324	0.00	-2.654
ZrO	650.5	653.0	646.2	647.2		833.191	4.38	-0.54		426.2327	0.00	-1.353
						810.832	2.72	-3.89		426.2346	0.00	-1.353
TiO	544.8	545.4	541.6	544.7	Y I	792.414	4.76	-1.68	Ce II	426.2348	0.00	-2.541
TiO	559.1	560.0	541.6	544.7		748.174	2.75	-4.10		426.2350	0.00	-1.549
TiO	561.5	562.0	580.0	581.0		746.152	2.55	-3.58		426.2364	0.00	-1.549
TiO	668.1	670.6	646.2	647.2		744.302	4.18	-1.82		426.2366	0.00	-1.590
TiO	671.4	673.5	646.2	647.2		743.053	2.58	-3.86		876.991	0.55	-2.37
TiO	705.4	706.9	703.0	705.0	Y II	735.350	4.73	-1.57		875.799	0.59	-2.29
TiO	712.5	714.4	703.0	705.0		880.059	0.00	-2.24		872.131	0.33	-2.48
Na	588.3	590.3	580.0	581.0		734.646	2.01	-0.82		870.237	0.47	-1.46
						788.188	1.84	-0.57		870.075	0.46	-2.77
						745.028	1.75	-1.59	La II	840.525	0.30	-2.10
					Zr I	784.937	0.69	-1.30		840.413	0.70	-1.67
						781.937	1.82	-0.38		839.451	0.27	-2.59
						411.689	0.00	-1.18		826.405	0.00	-2.56
						419.509	0.02	-0.91		826.342	0.33	-2.24
					Nb I	434.531	0.00	-1.36		824.047	0.32	-2.21
						424.946	0.00	-1.47		790.926	0.32	-2.43
						426.205	0.13	-0.56		789.897	0.90	-1.12
						535.072	0.27	-0.86		751.876	0.46	-2.94
						426.2215	0.00	-0.879		480.900	0.24	-1.40
					Tc I	426.2219	0.00	-1.527		480.404	0.24	-1.49
						426.2252	0.00	-1.527		474.873	0.93	-0.54
						426.2257	0.00	-1.191		466.249	0.00	-2.25
						426.2261	0.00	-1.337		455.846	0.32	-0.97
						426.2290	0.00	-1.337		433.375	0.17	-0.06
						426.2293	0.00	-1.655		432.246	0.17	-2.38

a. Boundaries (expressed in nm) of the continuum ($\lambda_{C,i}$, $\lambda_{C,f}$) and band ($\lambda_{B,i}$, $\lambda_{B,f}$) windows used in the computation of the band indices (equation (1)). **b.** Atomic line data (wavelength, excitation potential and oscillator strength²³) for the elements with abundances derived in the present study. For Zr, lines were restricted to those two lines having laboratory oscillator strengths³⁶. The hyperfine structure of Tc is included.

Extended Data Table 6 | Derived s-process timescales, counted from the first thermal pulse, and masses for the intrinsic S stars of our sample

Star name	M_{HR} (M_{\odot})	$\log(L/L_{\odot})$	C/O	R	Mass loss	t_{TP} (Myr)	$t_{\text{TP,min}}$ (Myr)	$t_{\text{TP,max}}$ (Myr)	TP	WP	M_{mod} (M_{\odot})	λ_{pm}	$\lambda_{\text{pm}}^{\text{min}}$	$\lambda_{\text{pm}}^{\text{max}}$
σ^1 Ori	2.5	3.45	0.50	0.049	VW	1.3	0.7	2.6	9	1.00	2	10%	2%	30%
					S	1.8	0.7	3.9	9	1.00	3	5%	2%	30%
AA Cam	3.0	3.91	0.50	0.055	VW	1.4	0.9	2.3	8	0.99	3	5%	2%	10%
					S	1.6	0.7	2.7	8	1.00	3	5%	2%	10%
KR CMa	–	–	0.50	0.052	VW	1.6	0.7	2.6	10	1.00	3	10%	2%	30%
					S	1.8	0.7	3.9	9	1.00	3	10%	2%	30%
CSS 454	–	–	0.50	0.055	VW	1.7	1.1	2.6	12	0.88	2	10%	5%	10%
					S	1.7	1.0	3.9	12	0.88	2	10%	5%	30%
HIP 103476	2.0	3.59	0.50	–	VW	2.2	1.2	3.2	14	1.00	3	10%	5%	10%
					S	2.5	1.5	4.0	13	1.00	3	10%	5%	10%
AD Cyg	–	–	0.97	0.072	VW	2.7	2.1	3.9	18	0.57	3	10%	3%	10%
					S	3.0	2.2	4.0	17	0.53	3	10%	3%	10%
NQ Pup	1.5	2.95	0.50	0.062	VW	2.8	2.1	4.2	19	0.28	3	3%	2%	3%
					S	3.1	2.3	4.0	18	0.33	3	10%	3%	10%
HR Peg	2.0	3.39	0.75	0.076	VW	3.2	2.4	4.8	23	0.94	3	3%	2%	5%
					S	3.0	2.4	3.7	17	0.73	3	3%	3%	10%

The adopted mass-loss prescription is indicated by VW¹⁹ or S²⁰. The TP column indicates the number of thermal pulses that took place in the star, up to the present time. The WP column lists the maximum of the weight product function (computed from equation (5)). M_{mod} and M_{HR} stand for the masses derived from nucleosynthesis models and stellar-evolution models (that is, location in the Hertzsprung–Russell diagram; Extended Data Fig. 3), respectively. The partial mixing parameter λ_{pm} expresses the ratio between the mass of the proton injection region and that of the thermal pulse. The C/O column provides an estimate of the C/O ratio (selected among the grid points C/O = 0.5, 0.75, 0.9, 0.92, 0.97, 0.99) best matching the spectroscopic and photometric data. The infrared excess $R = F(12\text{ }\mu\text{m})/F(2.2\text{ }\mu\text{m})$ is listed in the last column.

Optically addressable nuclear spins in a solid with a six-hour coherence time

Manjin Zhong¹, Morgan P. Hedges^{1,2}, Rose L. Ahlefeldt^{1,3}, John G. Bartholomew¹, Sarah E. Beavan^{1,4}, Sven M. Wittig^{1,5}, Jevon J. Longdell⁶ & Matthew J. Sellars¹

Space-like separation of entangled quantum states is a central concept in fundamental investigations of quantum mechanics and in quantum communication applications. Optical approaches are ubiquitous in the distribution of entanglement because entangled photons are easy to generate and transmit. However, extending this direct distribution beyond a range of a few hundred kilometres^{1,2} to a worldwide network is prohibited by losses associated with scattering, diffraction and absorption during transmission. A proposal to overcome this range limitation is the quantum repeater protocol^{3,4}, which involves the distribution of entangled pairs of optical modes among many quantum memories stationed along the transmission channel. To be effective, the memories must store the quantum information encoded on the optical modes for times that are long compared to the direct optical transmission time of the channel⁵. Here we measure a decoherence rate of 8×10^{-5} per second over 100 milliseconds, which is the time required for light transmission on a global scale. The measurements were performed on a ground-state hyperfine transition of europium ion dopants in yttrium orthosilicate ($^{151}\text{Eu}^{3+}:\text{Y}_2\text{SiO}_5$) using optically detected nuclear magnetic resonance techniques. The observed decoherence rate is at least an order of magnitude lower than that of any other system suitable for an optical quantum memory. Furthermore, by employing dynamic decoupling, a coherence time of 370 ± 60 minutes was achieved at 2 kelvin. It has been almost universally assumed that light is the best long-distance carrier for quantum information. However, the coherence time observed here is long enough that nuclear spins travelling at 9 kilometres per hour in a crystal would have a lower decoherence with distance than light in an optical fibre. This enables some very early approaches^{6,7} to entanglement distribution to be revisited, in particular those in which the spins are transported rather than the light.

Mapping optical quantum states onto the hyperfine states of rare-earth optical centres in crystals^{8–14} is an exciting avenue for ensemble-based quantum memories for light. Rare-earth-doped crystals possess optical transitions with homogeneous linewidths close to the lifetime limit¹⁵, high optical depths¹⁰, long-lived hyperfine states accessible via optical excitation¹⁶ and no spatial diffusion¹¹. There have been major advances in developing quantum memories based on rare-earth ions, including the demonstration of a gradient echo memory operating beyond the no-cloning limit¹⁰, and atomic frequency comb storage of single photons and quantum entanglement^{12,13}.

Recently, the reversible mapping of light onto the hyperfine states of europium-ion dopants in yttrium orthosilicate ($\text{Eu}^{3+}:\text{Y}_2\text{SiO}_5$) was demonstrated with sufficiently high efficiency and low enough noise floor for quantum memory applications^{17,18}. Nuclear spin states of $\text{Eu}^{3+}:\text{Y}_2\text{SiO}_5$ are particularly attractive for quantum protocols using spin-wave storage and manipulation due to their extremely long observed lifetime of 23 days and their optical accessibility¹⁶. Here we report a dramatic reduction in the decoherence of hyperfine transitions in $^{151}\text{Eu}^{3+}:\text{Y}_2\text{SiO}_5$

towards the lifetime limit, as well as a characterization of the remaining decoherence mechanisms.

The electronic ground-state ($^7\text{F}_0$) hyperfine structure of $^{151}\text{Eu}^{3+}$ ions occupying site 1 yttrium positions in Y_2SiO_5 (ref. 19) is shown in Fig. 1. At liquid helium temperatures, the dominant perturbation contributing to decoherence between these levels is the flipping of nearby yttrium (spin-1/2) nuclei causing a randomly fluctuating magnetic field at the Eu^{3+} site²⁰. Using the yttrium's gyromagnetic moment of 2.09 MHz T^{-1} and the procedure described in ref. 21, the degree of field fluctuation ΔB (full-width at half-maximum, FWHM) was calculated to be $8.0 \mu\text{T}$. Assuming that the fluctuations occur faster than the coherence time T_2 of the Eu^{3+} transition, T_2 can be estimated according to:

$$\frac{1}{\pi T_2} = S_1 \cdot \Delta B + \Delta B \cdot S_2 \cdot \Delta B \quad (1)$$

where S_1 and S_2 are the gradient (first-order term) and curvature (second-order term) of the transition frequency with respect to magnetic field²².

As shown in Fig. 1, level anti-crossings occur when the linear Zeeman splitting approaches the magnitude of the zero-field splitting due to the quadrupole interaction. In this regime, S_1 and S_2 are strongly dependent on the magnitude and direction of the applied field. By applying a particular field such that the transition frequency is at a critical point ($S_1 = 0$), the sensitivity of the hyperfine transition to ΔB can be reduced. This zero first-order Zeeman (ZEFOZ) method is analogous to a clock transition in atomic spectroscopy²³. At the ZEFOZ field, T_2 can be

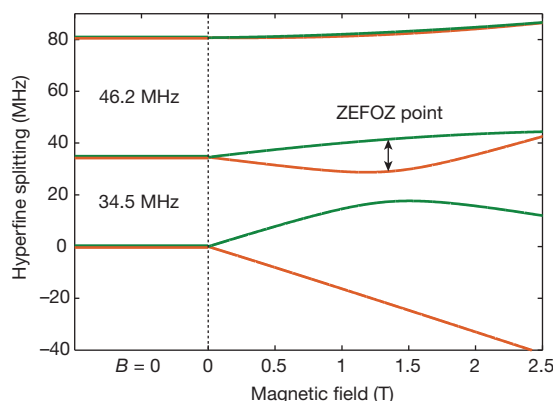


Figure 1 | Illustration of the ZEFOZ transition. Shown are the calculated hyperfine ground-state energy levels of $^{151}\text{Eu}^{3+}:\text{Y}_2\text{SiO}_5$ as the applied magnetic field is increased from zero in the direction of the targeted zero first-order Zeeman (ZEFOZ) transition. This calculation used the reduced spin Hamiltonian in ref. 24. The doubly degenerate hyperfine states ($I = \pm 1/2, \pm 3/2, \pm 5/2$) are split by the applied field. At a field magnitude of 1.35 T in the appropriate direction, the marked transition experiences a critical point. At this ZEFOZ point, the transition is, to first order, insensitive to small field perturbations in any direction (although only one direction is shown).

¹Centre for Quantum Computation and Communication Technology, Laser Physics Centre, The Australian National University, Canberra, Australian Capital Territory 0200, Australia. ²Department of Physics, Princeton University, Princeton, New Jersey 08554, USA. ³Laboratoire Aimé Cotton, CNRS-UPR 3321, Université Paris-Sud and ENS Cachan, 91405 Orsay, France. ⁴Fakultät für Physik und Center for NanoScience (CeNS), Ludwig-Maximilians-Universität, Geschwister-Scholl-Platz 1, 80539 Munich, Germany. ⁵Kayser-Threde GmbH, Wolfratshauser Straße 48, 81379 Munich, Germany. ⁶The Dodd-Walls Centre for Photonic and Quantum Technologies, and Department of Physics, University of Otago, 730 Cumberland Street, Dunedin 9016, New Zealand.

extended because only the second-order Zeeman coefficient S_2 , that is, the curvature at the critical point, couples the transition to magnetic fluctuations.

Previous studies of hyperfine transition decoherence in $\text{Pr}^{3+}:\text{Y}_2\text{SiO}_5$ indicate that the storage times achieved in this material are reaching the limit imposed by S_2 (refs 9, 11). Although a significant increase in current coherence times is unlikely in $\text{Pr}^{3+}:\text{Y}_2\text{SiO}_5$, it was predicted that much longer coherence times should be achievable in $\text{Eu}^{3+}:\text{Y}_2\text{SiO}_5$ using the ZEFOZ technique²⁴. This is because the second-order Zeeman coefficient is two orders of magnitude smaller in $\text{Eu}^{3+}:\text{Y}_2\text{SiO}_5$ than in $\text{Pr}^{3+}:\text{Y}_2\text{SiO}_5$.

We calculated the ZEFOZ magnetic fields for $^{151}\text{Eu}^{3+}:\text{Y}_2\text{SiO}_5$ using the reduced spin Hamiltonian determined in ref. 24. The investigation here focused on an easily accessible critical point of the $+3/2 \leftrightarrow -3/2$ transition that is predicted to occur at an applied field of 1.35 ± 0.07 T. At this field, the transition frequency is 12.45 MHz and the maximum curvature $|S_2| = 21.7 \text{ MHz T}^{-2}$. From equation (1), the expected decoherence rate is $4.4 \times 10^{-3} \text{ s}^{-1}$ for a perfectly aligned magnetic field. Using an iterative field optimization process (see ‘Magnetic field alignment’ in Methods), the first-order Zeeman coefficient $|S_1|$ was reduced below $1 \times 10^{-3} \text{ MHz T}^{-1}$, corresponding to a misalignment of less than 0.004° from the ideal critical point field orientation. This field alignment is labelled as \mathbf{B}_{cp} , for which a decoherence rate of $2.9 \times 10^{-2} \text{ s}^{-1}$ is expected according to equation (1).

Figure 2a illustrates the experimental set-up used to measure the decoherence rate of the $+3/2 \leftrightarrow -3/2$ transition. The measured decay of the two-pulse spin echo as a function of the delay τ is shown in Fig. 3a. Three data sets are presented, corresponding to the fields $|\mathbf{B}_{\text{cp}}|$, $(|\mathbf{B}_{\text{cp}}| + 0.005 \text{ T})$ and $(|\mathbf{B}_{\text{cp}}| + 0.05 \text{ T})$. These show a dramatic decrease in the echo decay rate as the field alignment brings the transition closer to the critical point. The echo attenuation envelope at \mathbf{B}_{cp} has a non-exponential form and cannot be properly described by a standard relaxation time T_2 because the decoherence rate changes with delay time τ . The decay is accurately described by the expression developed in ref. 25, which incorporates decoherence due to spectral diffusion:

$$E(\tau) \propto e^{-\left(\frac{\tau}{T_M}\right)^2} \quad (2)$$

where E is the echo amplitude as a function of τ . The phase memory time T_M is used to describe the coherence decay of the system. This was determined to be 47 s at \mathbf{B}_{cp} .

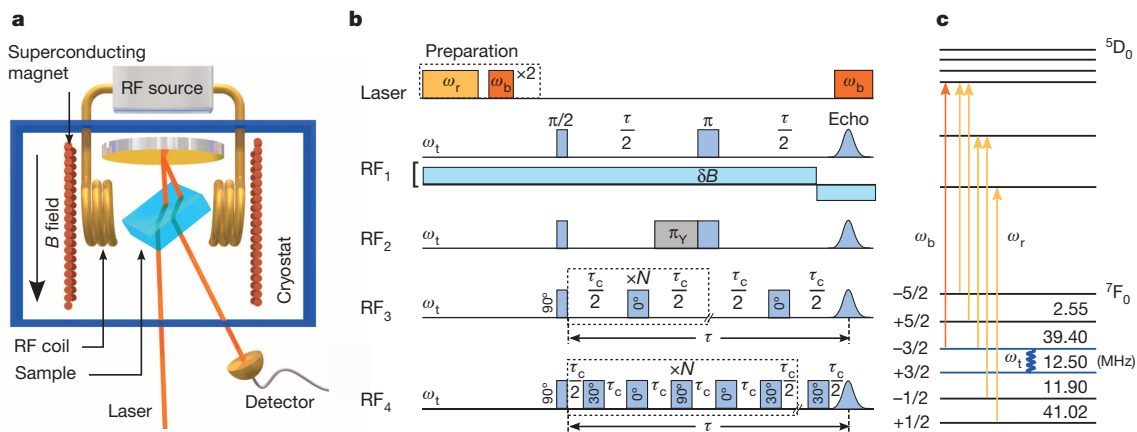


Figure 2 | Experimental set-up and pulse sequences. **a**, The $\text{Eu}^{3+}:\text{Y}_2\text{SiO}_5$ sample ($3 \times 3 \times 5 \text{ mm}$) is oriented within a superconducting magnet contained in a liquid helium cryostat. The RF (radio frequency) coil is aligned perpendicular to the magnet. The light from a highly stabilized dye laser (sub-kHz linewidth) is directed at the sample from below the magnet, reflected by a mirror and finally detected by a 125 MHz bandwidth Si photo-receiver (‘Detector’). **b**, The optical burn frequency ω_b is used to prepare a population difference between the $+3/2$ and $-3/2$ states before the application of RF pulses and optically detect the spin echo signal. The re-pump frequency

For $\tau \ll T_M$ the decoherence rate, given by $2\tau/T_M^2$ (the gradient of the echo decay), was extremely slow. For example, at $\tau = 100 \text{ ms}$, the fit to the model of ref. 25 predicts the decoherence rate to be $9.1 \times 10^{-5} \text{ s}^{-1}$, which is significantly slower than the rate of $4.4 \times 10^{-3} \text{ s}^{-1}$ for a perfect field alignment predicted by equation (1). This is attributed to the reduced local spin-reconfiguring rate due to the ‘frozen core’ effect²⁶ (see Fig. 3b). When a strong magnetic field is applied, a large magnetic moment is induced on each Eu ion’s electronic ground state, which is independent of the nuclear spin state. This creates a region (‘core’) around the Eu in which the mutual spin flips are suppressed because the Larmor frequencies of the Y spins are detuned from the bulk Y spins and from each other. For the experimentally determined ZEFOZ field of 1.29 T (see ‘Magnetic field alignment’ in Methods), and using the quadratic Zeeman parameters from ref. 27, we calculate the induced moment to be of the order of 100 MHz T^{-1} . The dipole–dipole interaction between the Eu^{3+} electronic state and the nearest neighbour Y spins is calculated to be 2 kHz, significantly larger than the 10 Hz interactions between the Y spins. This detuning reduces the spin flipping rate of these nearest neighbour Y spins. Hence, the spins in the frozen core that induce the largest field on the Eu site reconfigure at a suppressed rate, leading to a slow decoherence rate at short timescales.

We investigated the effect of the frozen core Y spins by intentionally ‘thawing’ the core during the echo using pulse sequence RF_2 (see Fig. 2 and ‘Investigation of the frozen core’ in Methods). The observed increase in decoherence rate is shown in Fig. 3a. For $\tau > 10 \text{ s}$, the frozen core has completely thawed and the Hahn echo decay is exponential. The decoherence rate of the exponential decay is $5.3 \times 10^{-2} \text{ s}^{-1}$, which is within a factor of two of the predicted decoherence rate from the directly measured transition sensitivity ($2.9 \times 10^{-2} \text{ s}^{-1}$). This confirms that the decoherence at \mathbf{B}_{cp} is still dominated by the interaction between the Eu^{3+} ions and the Y spin bath. Hence, the model of ref. 25 allows T_M to be related to the flipping rate R of the local Y spins and $\Delta\omega_{1/2}$, the half-width of the spectral broadening on the Eu hyperfine transition due to the distribution of the static local field:

$$T_M = 1.41 (R\Delta\omega_{1/2})^{-1/2} \quad (3)$$

When the frozen core has been removed, the decoherence rate ($5.3 \times 10^{-2} \text{ s}^{-1}$) is a measure of $\Delta\omega_{1/2}$. Therefore, from equation (3), the spin flip rate of the frozen core Y spins is $R = 1.7 \times 10^{-2} \text{ s}^{-1}$, which is four orders of magnitude slower than the zero-field value measured in ref. 20.

$\omega_r = [(\omega_b - 45 \text{ MHz}) \rightarrow \omega_b]$, scanning over 45 MHz, is used to pump ions back to the $+3/2$ state. Sequence RF_1 shows the RF pulses for phase sensitive detection with the polarity of a perturbing magnetic field reversed before the echo detection period; RF_2 is for measuring the frozen core Y³⁺ dynamics where a perturbing π_Y -pulse is applied just before the europium π -pulse; RF_3 is the CPMG pulse sequence; RF_4 is the KDD_x pulse sequence. **c**, The optical and RF frequencies used in **b**. The hyperfine transition frequencies as labelled are the theoretical values at the ZEFOZ field investigated.

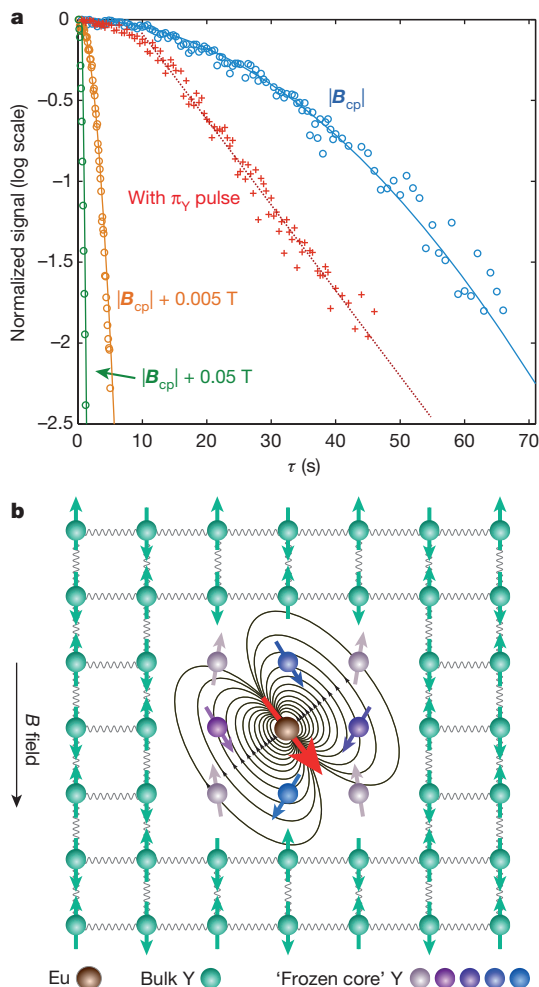


Figure 3 | Two-pulse spin echo decay and illustration of frozen core. **a**, Two-pulse spin echo decay with the echo amplitude E plotted against the total delay τ . Open circles show echo decays at magnetic fields $|B_{cp}|$ T (blue), $(|B_{cp}| + 0.005$ T; orange) and $(|B_{cp}| + 0.05$ T; green) as labelled. Each set is fitted using equation (2) with phase memory times T_M equal to 47 s, 3.5 s and 0.8 s respectively. 'Plus' symbols show the two-pulse spin echo decay with the perturbing π_V -pulse (sequence RF₂ in Fig. 2b). At $\tau > 10$ s, the decoherence rate is $5.3 \times 10^{-2} \text{ s}^{-1}$ from an exponential fit (that is, $E(\tau) \propto \exp(-\tau/T_2)$). **b**, Illustration of the 'frozen core' effect. A large dipole moment (red arrow) is induced on the Eu ion due to the large external field. The field due to the induced moment then alters the Larmor frequency and the quantization axis of nearby Y nuclei to create a 'frozen core' of yttriums that cannot readily exchange spin with the bulk Y nuclei. See key at bottom for details.

Owing to the very small amount of decoherence, the signal to noise ratio of the two-pulse spin echo measurements shown in Fig. 3a were not sufficient to accurately measure the decoherence rate for $\tau \ll T_M$. Therefore, a CPMG (Carr–Purcell–Meiboom–Gill) pulse sequence (RF₃ in Fig. 2) was used as a simple and effective method to measure the decoherence rate for $100 \text{ ms} < \tau < 10,000 \text{ ms}$. The results of the CPMG measurement are shown in Fig. 4. The echo decays exponentially as a function of the total evolution time, and the coherence time was increased as the time τ_c between the π -pulses was decreased. The decoherence rates measured by the CPMG pulse sequence are in good agreement with the gradient of the Hahn echo decay curve. For instance, the decoherence rate for CPMG at $\tau_c = 100 \text{ ms}$ is $8 \times 10^{-5} \text{ s}^{-1}$ ($T_2 = 200 \text{ min}$), which is consistent with the gradient of the Hahn echo decay at $\tau = 100 \text{ ms}$ in Fig. 3a ($9.1 \times 10^{-5} \text{ s}^{-1}$).

Although the CPMG sequence is commonly used to extend coherence, the method is only effective in preserving quantum states that possess a particular phase. Hence, it is not suitable for use in a practical quantum

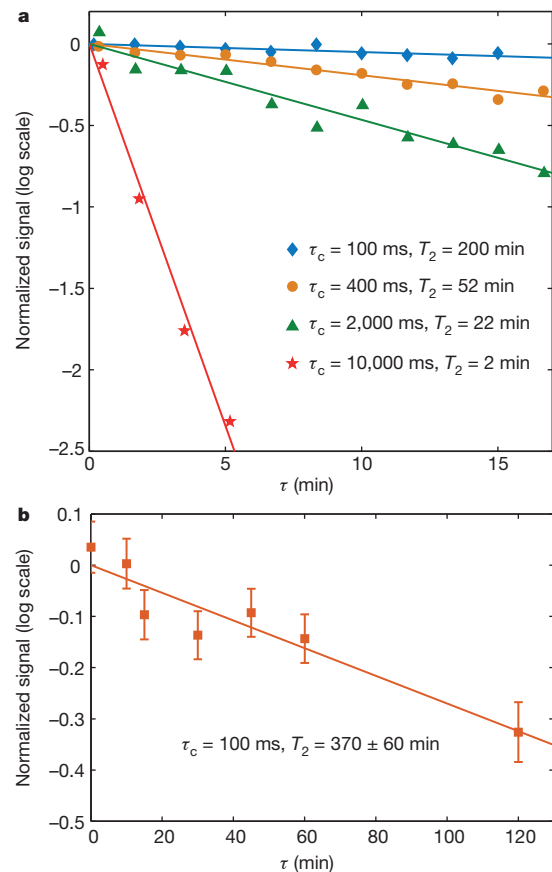


Figure 4 | Measured coherence times with CPMG and KDD_x pulse sequences. **a**, **b**, Echo decay at B_{cp} corresponding to CPMG pulse sequences (**a**) and KDD_x pulse sequences (**b**). The error bars in **b** correspond to the shot-to-shot variations in the echo amplitude due to the preparation sequence that were observed at short delays. The uncertainty in the calculated T_2 is obtained from the uncertainty in the linear fit.

memory, which is required to store arbitrary quantum states. As an example, the coherence time of 200 min achieved with CPMG was observed to decrease to less than 10 s when the phase of the initial $\pi/2$ -pulse was rotated by 90° . To test the suitability of the $\text{Eu}^{3+}:\text{Y}_2\text{SiO}_5$ system for arbitrary quantum state storage, a further experiment was performed using a dynamic decoupling (DD) sequence known as KDD_x. The KDD_x sequence (RF₄ in Fig. 2) is effective in preserving arbitrary states, and is robust against errors due to pulse area and off-resonant excitation²⁸. As shown in Fig. 4, a coherence time of $370 \pm 60 \text{ min}$ (6 ± 1 hours) was observed for a decoupling period τ_c of 100 ms. This slightly longer coherence time compared to the result of the CPMG experiment is attributed to the day-to-day alignment variation of the magnetic field on the order of 0.001° .

Equation (1) predicts that the coherence time could be improved by an order of magnitude by improving the alignment of the applied field such that $|S_1| = 0$. However, the current alignment is already close to the limit imposed by the ensemble inhomogeneity. Each ion possesses a slightly different critical point magnetic field due to inhomogeneity in the quadrupole interaction throughout the crystal. Thus, every ion cannot be maximally decoupled from magnetic interactions simultaneously. Despite this limitation, even with the current alignment precision, the coherence time could be extended by moving to a critical point with a lower $|S_2|$. There are many critical points that occur for ground-state transitions in $^{151}\text{Eu}^{3+}:\text{Y}_2\text{SiO}_5$, and a number of these have slightly smaller values of $|S_2|$ than the point studied in this Letter. In addition, the $|S_2|$ values for critical points in $^{153}\text{Eu}^{3+}:\text{Y}_2\text{SiO}_5$, the other naturally occurring Eu isotope, are smaller still by approximately a factor of two. Furthermore, many of these critical points occur at magnetic field values

up to five times larger than the field applied in this work. By increasing the magnetic field, the increased magnetic moment induced on the europium ion would extend the volume of the frozen core. The direct benefit would be a reduction of the Y spin flip rate R , which significantly reduces the rate at which DD rephasing pulses would need to be applied. While the bulk Y spin flips remain the dominant decoherence mechanism, these refinements of the ZEFOZ technique would allow advances towards the lifetime limit (see 'Hyperfine state lifetime at the ZEFOZ field' in Methods).

The preservation of quantum coherence on an hour-long timescale achieved via the KDD_x sequence demonstrates that it is possible to extend $\text{Eu}^{3+}:\text{Y}_2\text{SiO}_5$ quantum memories to long-term storage of arbitrary quantum states. Similar storage times have recently been observed²⁹ for the nuclear spin transition of ionized $^{31}\text{P}^+:\text{Y}_2\text{SiO}_5$ dopants in isotopically pure ^{28}Si . However, there are important distinctions between the $^{31}\text{P}^+:\text{Y}_2\text{SiO}_5$ system and the rare-earth-ion-doped crystal investigated in this Letter. The first is the ability to optically address the storage transition. The mapping of optical quantum states onto the hyperfine transitions of $^{151}\text{Eu}^{3+}:\text{Y}_2\text{SiO}_5$ has been previously demonstrated¹⁴, whereas a transition suitable for mapping light onto the spin transition of interest in $^{31}\text{P}^+:\text{Y}_2\text{SiO}_5$ has not been identified. The second distinction is the method used to decouple the storage transition from environmental perturbations. The long coherence times achieved in ref. 29 were possible because the dominant source of magnetic perturbations, the ^{29}Si spins, were removed. By contrast, in the current work the ZEFOZ technique was used to dramatically reduce the transition's sensitivity to magnetic field fluctuations. As a result, long coherence times were achieved despite the crystal being stoichiometric in perturbing spins and the cryostat being without any additional magnetic shielding.

The extension of this work to storage and retrieval of optical states would open up the possibility of remote entanglement distribution via the physical transport of memory crystals. A particularly attractive aspect of this idea is that it would allow entanglement distribution between remote locations with minimal permanent infrastructure. The insensitivity of the transition to external magnetic fields indicates that straightforward magnetic shielding should allow these properties to be preserved in transit. Rare-earth-ion quantum memories already boast high recall efficiencies¹⁰ and large data storage capacity³⁰. Incorporating such properties with long-term storage has the potential for a more efficient method of transporting quantum information. At a transport speed of 9 km h^{-1} the losses during transportation would be $0.16 \pm 0.03\text{ dB km}^{-1}$, which is lower than photon transmission losses in optical fibres. Even at the moderate transport speed of 100 km h^{-1} , storage in a memory could preserve $\sim 14\%$ of the input photons over a distance of 600 km, whereas the fraction of photons transmitted through this length of fibre is of the order of 10^{-12} . The scale of entanglement separation that should be possible with this approach presents a new regime for both quantum communication and fundamental tests of quantum mechanics.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 25 July; accepted 28 October 2014.

1. Dynes, J. F. *et al.* Efficient entanglement distribution over 200 kilometers. *Opt. Express* **17**, 11440–11449 (2009).
2. Yin, J. *et al.* Quantum teleportation and entanglement distribution over 100-kilometre free-space channels. *Nature* **488**, 185–188 (2012).
3. Briegel, H. J., Dür, W., Cirac, J. I. & Zoller, P. Quantum repeaters: the role of imperfect local operations in quantum communication. *Phys. Rev. Lett.* **81**, 5932–5935 (1998).
4. Duan, L. M., Lukin, M. D., Cirac, J. I. & Zoller, P. Long-distance quantum communication with atomic ensembles and linear optics. *Nature* **414**, 413–418 (2001).
5. Sangouard, N., Simon, C., de Riedmatten, H. & Gisin, N. Quantum repeaters based on atomic ensembles and linear optics. *Rev. Mod. Phys.* **83**, 33–80 (2011).

6. Bohm, D. Discussion of experimental proof for the paradox of Einstein, Rosen, and Podolsky. *Phys. Rev.* **108**, 1070–1076 (1957).
7. Bennett, C. H., Brassard, G., Breidbart, S. & Wiesner, S. in *Advances in Cryptology: Proceedings of CRYPTO '82* (eds Chaum, D., Rivest, R. L. & Sherman, A. T.) 267–275 (Plenum, 1982).
8. Longdell, J. J., Fraval, E., Sellars, M. J. & Manson, N. B. Stopped light with storage times greater than one second using electromagnetically induced transparency in a solid. *Phys. Rev. Lett.* **95**, 063601 (2005).
9. Fraval, E., Sellars, M. J. & Longdell, J. J. Dynamic decoherence control of a solid-state nuclear-quadrupole qubit. *Phys. Rev. Lett.* **95**, 030506 (2005).
10. Hedges, M. P., Longdell, J. J., Li, Y. & Sellars, M. J. Efficient quantum memory for light. *Nature* **465**, 1052–1056 (2010).
11. Heinze, G., Hubrich, C. & Halfmann, T. Stopped light and image storage by electromagnetically induced transparency up to the regime of one minute. *Phys. Rev. Lett.* **111**, 033601 (2013).
12. Clausen, C. *et al.* Quantum storage of photonic entanglement in a crystal. *Nature* **469**, 508–511 (2011).
13. Saglamyurek, E. *et al.* Broadband waveguide quantum memory for entangled photons. *Nature* **469**, 512–515 (2011).
14. Timoney, N., Lauritzen, B., Usmani, I., Afzelius, M. & Gisin, N. Atomic frequency comb memory with spin-wave storage in $^{153}\text{Eu}^{3+}:\text{Y}_2\text{SiO}_5$. *J. Phys. B* **45**, 124001 (2012).
15. Thiel, C. W., Böttger, T. & Cone, R. L. Rare-earth-doped materials for applications in quantum information storage and signal processing. *J. Lumin.* **131**, 353–361 (2011).
16. Könz, F. *et al.* Temperature and concentration dependence of optical dephasing, spectral-hole lifetime, and anisotropic absorption in $\text{Eu}^{3+}:\text{Y}_2\text{SiO}_5$. *Phys. Rev. B* **68**, 085109 (2003).
17. Timoney, N., Usmani, I., Jobez, P., Afzelius, M. & Gisin, N. Single-photon-level optical storage in a solid-state spin-wave memory. *Phys. Rev. A* **88**, 022324 (2013).
18. Jobez, P. *et al.* Cavity-enhanced storage in an optical spin-wave memory. *New J. Phys.* **16**, 083005 (2014).
19. Yano, R., Mitsunaga, M. & Uesugi, N. Ultralong optical dephasing time in $\text{Eu}^{3+}:\text{Y}_2\text{SiO}_5$. *Opt. Lett.* **16**, 1884–1886 (1991).
20. Arcangel, A., Lovrić, M., Tumino, B., Ferrier, A. & Goldner, P. Spectroscopy and coherence lifetime extension of hyperfine transitions in $^{153}\text{Eu}^{3+}:\text{Y}_2\text{SiO}_5$. *Phys. Rev. B* **89**, 184305 (2014).
21. Fraval, E. *Minimising the Decoherence of Rare Earth Ion Solid State Spin Qubits*. Ph.D. thesis, Australian National Univ. (2005).
22. Fraval, E., Sellars, M. J. & Longdell, J. J. Method of extending hyperfine coherence times in $\text{Pr}^{3+}:\text{Y}_2\text{SiO}_5$. *Phys. Rev. Lett.* **92**, 077601 (2004).
23. Langer, C. *et al.* Long-lived qubit memory using atomic ions. *Phys. Rev. Lett.* **95**, 060502 (2005).
24. Longdell, J. J., Alexander, A. L. & Sellars, M. J. Characterization of the hyperfine interaction in europium-doped yttrium orthosilicate and europium chloride hexahydrate. *Phys. Rev. B* **74**, 195101 (2006).
25. Mims, W. B. Phase memory in electron spin echoes, lattice relaxation effects in $\text{CaWO}_4:\text{Er}$, Ce, Mn. *Phys. Rev.* **168**, 370–389 (1968).
26. Shelby, R. M., Yannoni, C. S. & Macfarlane, R. M. Optically detected coherent transients in nuclear hyperfine levels. *Phys. Rev. Lett.* **41**, 1739–1742 (1978).
27. Thorpe, M. J., Leibbrandt, D. R. & Rosenband, T. Shifts of optical frequency references based on spectral-hole burning in $\text{Eu}^{3+}:\text{Y}_2\text{SiO}_5$. *New J. Phys.* **15**, 033006 (2013).
28. Souza, A. M., Álvarez, G. A. & Suter, D. Robust dynamical decoupling for quantum computing and quantum memory. *Phys. Rev. Lett.* **106**, 240501 (2011).
29. Saeedi, K. *et al.* Room-temperature quantum bit storage exceeding 39 minutes using ionized donors in silicon-28. *Science* **342**, 830–833 (2013).
30. Bonarota, M., Le Gouët, J.-L. & Chanelière, T. Highly multimode storage in a crystal. *New J. Phys.* **13**, 013013 (2011).

Acknowledgements We thank N. Manson and S. Rogge for comments on the manuscript. This work was supported by the Australian Research Council Centre of Excellence for Quantum Computation and Communication Technology (CE110001027), and M.J.S. was supported by an Australian Research Council Future Fellowship (FT110100919). J.J.L. was supported by the Marsden Fund of the Royal Society of New Zealand (contract U001221).

Author Contributions The initial project was conceived by M.J.S. and J.J.L. The experimental apparatus and initial experiments were designed and implemented by M.Z. assisted by M.P.H., S.M.W., S.E.B. and M.J.S. The precision alignment of the sample, coherence measurements and yttrium study were performed by M.Z. supported by R.L.A., J.G.B. and M.J.S. Analysis and interpretation of the data was performed by M.Z. and M.J.S. in consultation with R.L.A. and J.G.B. Modelling of the yttrium spin bath and europium hyperfine Hamiltonian was completed by M.Z. with assistance from J.J.L., R.L.A. and M.J.S. The paper was written by M.Z., J.G.B. and M.J.S. in discussion with all remaining authors.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to M.Z. (grace.zhong@anu.edu.au).

METHODS

Experimental set-up. As shown in Fig. 2a, a 0.01% doped $\text{Eu}^{3+}:\text{Y}_2\text{SiO}_5$ crystal from Scientific Materials Corp. (Bozeman, Montana) was maintained at 2 K in a bath cryostat that incorporates a superconducting magnet. The hyperfine transition was excited by a six turn RF coil with a diameter of 6 mm. The RF coil was driven by a 40 dBm RF source, which resulted in a Rabi frequency $\Omega_{\text{RF}} = 7.9$ kHz. Raman heterodyne detection was used to observe spin coherence optically via the ${}^7\text{F}_0 \leftrightarrow {}^5\text{D}_0$ transition at 579.87985 nm (air)¹⁹. The laser power incident on the crystal was 30 mW and was gated by an 80 MHz acousto-optic modulator (AOM).

The optical and RF pulse sequences used in the spin echo experiments are shown in Fig. 2b. The system was prepared with an optical re-pump and burn scheme to initialize the transition of interest. First, the laser frequency was swept over 45 MHz on the low frequency side of the $-3/2 ({}^7\text{F}_0) \leftrightarrow +3/2 ({}^5\text{D}_0)$ transition for 2 s (ω_r). The effect of the sweep was to optically re-pump ions in this subgroup to the $+3/2$ ground state, while ions in other subgroups were shelved to the $\pm 1/2$ ground states. A further pulse resonant with the $-3/2 ({}^7\text{F}_0) \leftrightarrow +3/2 ({}^5\text{D}_0)$ transition (ω_b) ensured that the $-3/2$ ground state was emptied. Both the re-pump (ω_r) and burn (ω_b) were repeated a second time to achieve the desired population distribution amongst the ground-state hyperfine levels.

Once the ensemble was prepared in the $+3/2$ ground state, an RF pulse sequence ($\text{RF}_{1,2,3,4}$) was applied depending on the measurement of interest. In all cases the resultant spin coherence was read out with Raman heterodyne detection by applying a further optical pulse at ω_b at the time of rephasing.

This preparation scheme was used because it allowed the transition of interest to be initialized independent of the magnitude and direction of the applied magnetic field close to the ZEFOZ field. For quantum memory applications where the magnetic field is fixed an optical pumping scheme tailored to specific optical transition frequencies, similar to those used in ref. 10 would be equally suitable.

Magnetic field alignment. The alignment of the sample with respect to the magnetic field direction was achieved by minimising the sensitivity (that is, S_1) of the $+3/2 \leftrightarrow -3/2$ hyperfine transition to externally applied magnetic perturbations

δB . These perturbations, of the order of 1 mT, were applied along three orthogonal directions. The transition sensitivity was measured by observing the phase shift of a two-pulse Hahn echo that occurred when the perturbation field was reversed after the rephasing π -pulse (see sequence RF_1 in Fig. 2b). This phase shift is observed by mixing the beat signal of the Hahn echo with two separate local oscillator channels separated in phase by 90° . The crystal was orientated relative to a large static field using two piezo-driven goniometers and the magnitude of the static field was also adjusted. By iteratively minimizing the phase shift the optimum alignment was achieved at a field magnitude of 1.29 ± 0.01 T within 3° of the direction $[-0.535, -0.634, 0.558]$ in the $[D_1, D_2, b]$ coordinate frame of the crystal (where the b axis is the twofold symmetry axis of the crystal and the D_1 and D_2 axes are the other two principle axes of polarization).

Investigation of the frozen core. To intentionally thaw the frozen core, a π -pulse resonant with the Y Zeeman splitting at 2.70 MHz (π_Y -pulse) was applied just before the Eu π -pulse (pulse sequence RF_2 in Fig. 2b). The π_Y -pulse aims to invert the Y spins to maximize their contribution to the decoherence rate. If the inversion of the frozen core was complete, the echo decay should become exponential. The non-exponential decay observed in Fig. 3a for $\tau < 10$ s is attributed to the fact that the π_Y -pulse possessed an insufficient bandwidth (200 Hz) to achieve total inversion and hence, could not maximally perturb the Eu^{3+} ions. Rather, the perturbing pulses were tuned to be resonant with the frozen core Y spins with the lowest flipping rate.

Hyperfine state lifetime at the ZEFOZ field. An initial investigation into the hyperfine state lifetime at the ZEFOZ field was performed by writing a spectral grating using repeated $\pi/2$ -pulse pairs then probing the grating with a third $\pi/2$ -pulse after a long delay. The decay of the resultant output echo amplitude as the delay between writing and probing the grating is increased provided a measure of the spin state lifetime. The results of the measurement confirmed that the coherence time was far from being limited by lifetime relaxation. Due to a large uncertainty because of short delay times (< 1 hour) arising from experimental limitations, such as the cryostat hold time, it was difficult to determine an accurate lifetime value. However, a minimum bound on the lifetime was determined to be 15 h.

High-value alcohols and higher-oxidation-state compounds by catalytic Z-selective cross-metathesis

Ming Joo Koh¹, R. Kashif M. Khan¹, Sebastian Torker¹, Miao Yu¹, Malte S. Mikus¹ & Amir H. Hoveyda¹

Olefin metathesis catalysts provide access to molecules that are indispensable to physicians and researchers in the life sciences^{1,2}. A persisting problem, however, is the dearth of chemical transformations that directly generate acyclic Z allylic alcohols, including products that contain a hindered neighbouring substituent or reactive functional units such as a phenol, an aldehyde, or a carboxylic acid. Here we present an electronically modified ruthenium–disulfide catalyst that is effective in generating such high-value compounds by cross-metathesis. The ruthenium complex is prepared from a commercially available precursor and an easily generated air-stable zinc catechthiolate. Transformations typically proceed with 5.0 mole per cent of the complex and an inexpensive reaction partner in 4–8 hours under ambient conditions; products are obtained in up to 80 per cent yield and 98:2 Z:E diastereoselectivity. The use of this catalyst is demonstrated in the synthesis of the naturally occurring anti-tumour agent neopeltolide and in a single-step stereoselective gram-scale conversion of a renewable feedstock (oleic acid) to an anti-fungal agent. In this conversion, the new catalyst promotes cross-metathesis more efficiently than the commonly used dichloro-ruthenium complexes, indicating that its utility may extend beyond Z-selective processes.

In 2009, it was demonstrated that a molybdenum (Mo) monopyrrolide aryloxide complex can promote ring-opening/cross-metathesis with high Z selectivity³. This initial discovery and the attendant stereochemical model paved the way for development of other Mo and tungsten (W) alkylidenes that facilitate Z-selective cross-metathesis⁴ and macrocyclic ring-closing metathesis⁵. It was later shown that olefin metathesis can be promoted Z selectively by ruthenium (Ru) complexes containing a bidentate N-heterocyclic carbene (NHC) with a carbanionic arm (such as **Ru-2**, Fig. 1a)^{6–8}. Monosulfide/monochloride Ru carbenes have also been used for Z-selective homocoupling of unhindered terminal alkenes, but Z:E ratios (where E is the other isomer) generally plummet at later stages of the transformation owing to post-olefin metathesis isomerization⁹. The most recently developed Ru complexes for Z-selective olefin metathesis, represented by **Ru-3a**, feature a catechthiolate ligand¹⁰; the utility of these carbenes has been confined to ring-opening metathesis polymerization¹⁰ and ring-opening/cross-metathesis¹¹, processes that benefit from the driving force of strain release.

That different families of complexes have been introduced for Z-selective catalysts is not surprising, since the range of functional group compatibility offered by Mo-, W- and Ru-based catalysts is a key reason for the rapid rise of olefin metathesis^{1,12}. Ru dichloride species (such as **Ru-1**) retain their structural integrity in the presence of a hydroxyl, an aldehyde and a carboxylic acid group, but Mo and W alkylidenes are rendered inactive under such circumstances, necessitating protection/deprotection schemes¹³. Unlike high-oxidation-state alkylidenes, Ru complexes are often deactivated in the presence of amines⁵ or enol ethers⁴. An NHC–Mo complex with a triflate ligand has been shown to tolerate alcohol, aldehyde and carboxylic acid functions, but only for extraordinarily facile ring-opening metathesis polymerization and ring-closing metathesis reactions¹⁴.

A readily accessible and easy-to-use Ru-based cross-metathesis catalyst that generates linear molecules containing an allylic alcohol¹⁵, a phenol, an aldehyde or a carboxylic acid (compare **i–iii**, Fig. 1b) would have a substantial impact on our ability to access a large assortment of important molecules; these functional units commonly occur in nature and are widely employed in chemistry. It is, however, not a given that the Z-selective carbene derivatives would exhibit the same degree of robustness and chemoselectivity as the dichloro-Ru systems. This is because the development of Z-selective Ru catalysts has been achieved through substitution of the halogens with other anionic ligands (Fig. 1a). In **Ru-2**, the chlorine atoms in a species such as **Ru-1** are replaced by an alkyl and a nitrate group, and in **Ru-3a** we have two Ru–S bonds instead. Anionic ligands in a Ru carbene catalyst certainly have a strong influence on its performance¹⁶. For example, although allylic alcohols are prone to competitive olefin isomerization¹⁷, they can undergo unusually efficient and selective olefin metathesis with Ru-dichloride complexes (such as **Ru-1**) partly owing to H-bonding with the chloride ligands¹⁸; such a feature could be lost as a result of alterations implemented in the design of Z-selective catalysts. Carbenes derived from **Ru-2** catalyse olefin metathesis in the presence of a remote alcohol but yields are at times diminished^{7,19}, and the related Z-selective ring-opening/cross-metathesis with allylic alcohols has involved only highly strained cyclobutene²⁰. Similarly, ring-opening/cross-metathesis with dithiolate **Ru-3a** and an allylic alcohol is limited to strained cyclic olefins¹¹. In the latter cases, as with the NHC-containing Mo alkylidenes¹⁴, rapid transformation with an exceedingly reactive cyclic alkene allows productive olefin metathesis to compete effectively with catalyst decomposition.

A pressing question then is whether any existing Ru catalyst can efficiently promote cross-metathesis reactions that generate the types of Z alkenes represented by **i–iii** in Fig. 1b. Z-selective cross-metathesis transformations that directly afford allylic alcohols are yet to be disclosed. Catalytic Z-selective olefin metathesis involving an aldehyde-containing or a carboxylic-acid-containing substrate is unknown (see **ii–iii**). Furthermore, unlike dichloro-Ru carbenes, complexes such as **Ru-2** are generally used with linear, unhindered aliphatic alkenes; with sterically hindered reaction partners, olefin metathesis is substantially less efficient⁶ (such as with a styrene; see **i**).

We selected Z-2-butene-1,4-diol (**1**; Fig. 1c) as the cross partner for two reasons: stereoisomerically pure **1** is commercially and inexpensively available, and use of 1,2-disubstituted olefin diminishes the amount of ethylene formation, discouraging the generation of unstable methylenes that are adept at promoting post-cross-metathesis isomerization⁴. In the event, cross-metathesis of **1** and allylbenzene (**2**) with 3.0 mol% **Ru-2** gave a Z:E = 91:9 diastereoisomeric mixture for product **3** in 50% yield (22 °C, 9 h); reaction with dithiolate **Ru-3a** was less efficient (42% yield) but more stereoselective (98:2 Z:E). Longer reaction times did not lead to further conversion, pointing to catalyst instability as a culprit. We surmised that a decomposition pathway could involve a sulfide shift (**A** → **B**, Fig. 1d), leading to generation of a catalytically inactive Ru-alkyl species. Catechthiolate complexes (such as **Ru-3a**) contain a sulfide oriented opposite to the NHC, a disposition designed to promote

¹Department of Chemistry, Merkert Chemistry Center, Boston College, Chestnut Hill, Massachusetts 02467, USA.

Table 1 | Cross-metathesis of **1** and **2** with modified complexes

Entry	Ru complex	Reaction duration (h)	Conversion (%)	Yield of 3 (%)	Z:E
1	Ru-3b ; 3.0 mol%	9	77	61	96:4
2	Ru-3c ; 3.0 mol%	9	75	60	94:6
3	Ru-3d* ; 3.0 mol%	9	76	57	95:5
4	Ru-3e ; 3.0 mol%	9	75	58	96:4
5	Ru-3f ; 3.0 mol%	9	56	47	96:4
6	Ru-3b ; 5.0 mol%	4	84	71	96:4

*Inseparable mixture of Ru complex isomers.

Reactions were carried out under a N₂ atmosphere. The percentage conversion was determined by analysis of the 400 MHz ¹H NMR spectra of the unpurified mixtures and the variance of the conversion values is estimated to be less than ±2%. The variance of the yield for the isolated and purified product **3** (Z:E mixture) is estimated to be less than ±5%. Z:E ratios were determined by ¹H NMR analysis of unpurified mixtures and the variance of the ratios is estimated to be less than ±2%. See Supplementary Information for all details.

theory) investigating the following processes (Fig. 1e): nonproductive olefin metathesis involving Z-2-butene (Bu), and competitive decomposition that entails the 1,2-shift of the more electron-rich *trans*-to-NHC sulfide ligand to the electrophilic ethylidene (compare Figure 1d; for details see Supplementary Information). The turnover-limiting barrier (transition state derived from Z-2-butene and the chloro complex **Ru-3b** **ts**_{Bu,Cl}) for the olefin metathesis reaction of the active 14-electron

dichlorocatechothiolate species (**14e**_{Cl}) with Z-2-butene to generate the metallacyclobutane structure (labelled as **mcb**_{Bu,Cl}; Fig. 1e, right) was calculated to be 2.1 kcal mol^{−1} lower in energy compared to **ts**_{Bu,H} (14.2 kcal mol^{−1} versus 12.1 kcal mol^{−1}; red versus blue curve); this points to increased Ru Lewis acidity in **14e**_{Cl} and a less endergonic olefin coordination (6.9 kcal mol^{−1} versus 8.8 kcal mol^{−1} for the π complexes **pc**_{Bu,Cl} versus **pc**_{Bu,H}). Additionally, the more stabilized metallacyclobutane

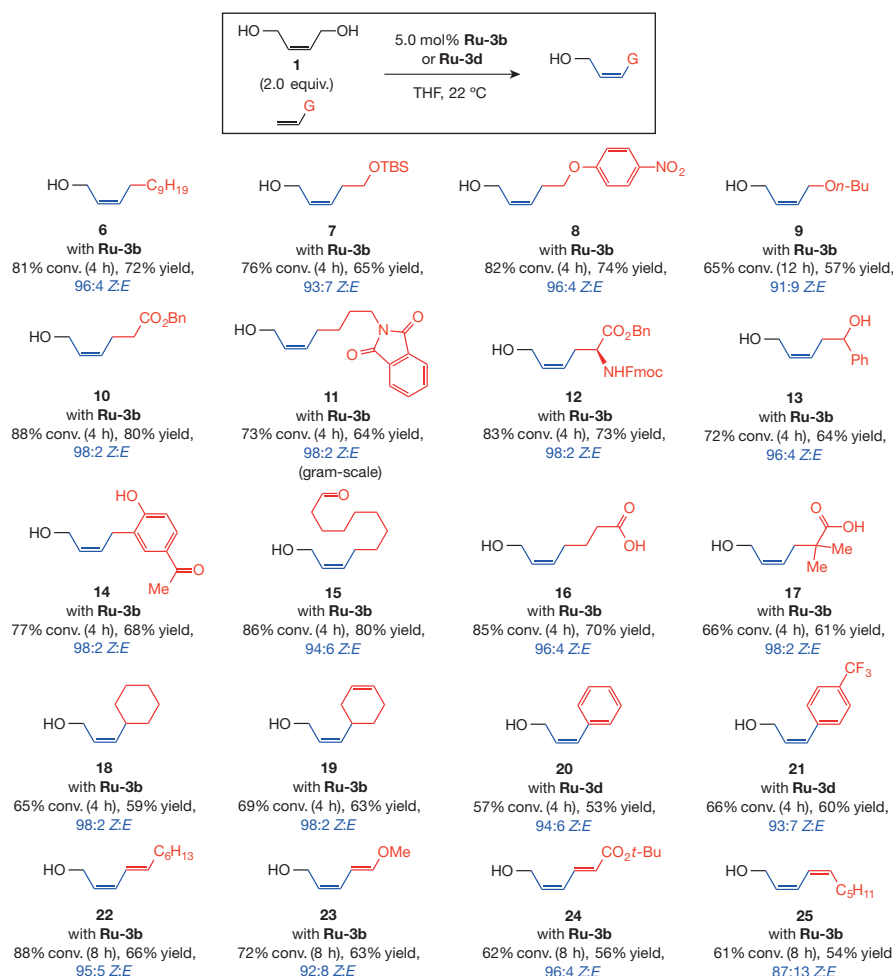


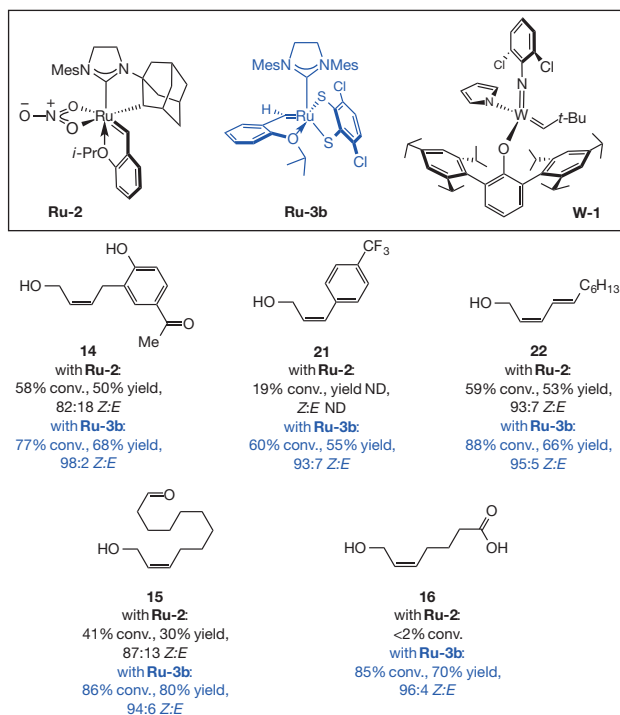
Figure 2 | The range of Z allylic alcohols made available through catalytic stereoselective cross-metathesis with **Ru-3b**. An assortment of Z allylic alcohols can be accessed through cross-metathesis reactions performed with 5.0 mol% **Ru-3b**. Transformations afford the desired products in 53%–80% yield and 87:13–98:2 Z:E. Cross-metathesis can occur between **1** and terminal alkenes that contain another hydroxyl group (**13**), a ketone (**14**), an aldehyde (**15**), or a carboxylic acid (**16** and **17**). Sterically hindered alkenes (**18**–**21**), Z- or E-1,3-dienes including those that bear an electron-releasing (**23**) or an electron-withdrawing (**24**) substituent can be used as well. For synthesis of Z-alkenes **20** and **21** the use of complex **Ru-3d** led to higher yields (44% yield for **20** and 55% yield for **21** with **Ru-3b**, and identical Z:E ratios (93:7–94:6 Z:E) in all cases). Conversions and stereoselectivities were determined by analysis of ¹H NMR spectra of the unpurified mixtures; yields correspond to isolated and purified products; 5.0 equivalents of **1** was used for **18**–**21**. See the Supplementary Information for reaction conditions and other details. Bn, benzyl; Fmoc, fluorenylmethoxycarbonyl; TBS, *tert*-butyldimethylsilyl.

derived from the dichlorocatechothiolate species hints at a weakened *trans* influence ($10.0 \text{ kcal mol}^{-1}$ versus $12.1 \text{ kcal mol}^{-1}$ for **mcb**_{Bu,Cl} versus **mcb**_{Bu,H}). Equally noteworthy is that, whereas generation of the propene-bound product resulting from a 1,2-shift within the modified complex (labelled as **shift prod**_{Pr,Cl}) from the corresponding propene-bound π complex (**pc**_{Pr,Cl}) is endergonic by $0.5 \text{ kcal mol}^{-1}$, the same process in the parent system (that is, **pc**_{Pr,H} \rightarrow **shift prod**_{Pr,H}) is appreciably exergonic ($-2.5 \text{ kcal mol}^{-1}$). We attribute the latter discrepancy to the alleviation of a stronger electronic repulsion between the Lewis basic NHC and non-chlorinated sulfide ligands upon cleavage of the Ru–S bond; Ru complexes **shift prod**_{Pr} are then generated that can further react through as yet unidentified mechanisms. The appreciably lower barrier ($9.5 \text{ kcal mol}^{-1}$) for the 1,2-shift induced by the π -acidic ethylene (structure of the complex not shown) suggests a detrimental effect in cross-metathesis between two terminal olefins versus when diol **1** is used. Overall, comparison of the energy barriers for the model olefin metathesis process involving Ru-ethylidene and Z-2-butene (proceeding through a transition state labelled as **ts**_{Bu}) versus the 1,2-shift induced by the propene (via a transition state labelled as **shift ts**_{Pr}) implies a favourable partitioning between productive and decomposition pathways in the case of **Ru-3b** ($12.1 \text{ kcal mol}^{-1}$ and $13.4 \text{ kcal mol}^{-1}$ for the transition states derived from olefin metathesis with Z-2-butene (**ts**_{Bu,Cl}) versus the 1,2-shift in the presence of bound propene (**shift ts**_{Pr,Cl}), respectively); for **Ru-3a**, the alternatives are equally preferred ($14.2 \text{ kcal mol}^{-1}$ for **ts**_{Bu,H} and **shift ts**_{Pr,H}). The halogen substituents thus lower for the π complex formed between Z-2-butene and the dichlorocatechothiolate system (**pc**_{Bu,Cl}) versus that which is derived from the catechothiolate species (**pc**_{Bu,H}).

Efforts to prepare **Ru-3b** (Fig. 1f) via sodium dithiolate precursors¹⁰ initially proved unsatisfactory, as the desired Ru catechothiolate was inseparable from unidentified side products ($>98\%$ conv.; $<10\%$ yield). We subsequently established that reaction of 3,6-dichloro-catechothiol **4** with ethylenediamine and $\text{Zn}(\text{OAc})_2 \cdot 2\text{H}_2\text{O}$ —all of which are commercially available—at room temperature for one hour affords the air-stable white solid Zn-dithiolate **5** (95% yield). Treatment of **5** with purchasable **Ru-1** gave **Ru-3b** in 85% yield with little or no byproduct formation, rendering purification of the derived Ru complex straightforward. Carbenes **Ru-3c** to **Ru-3f** (Table 1) were synthesized similarly. In support of the suggested effect of electronically altered catechothiolate ligands, the X-ray structure of **Ru-3b** (Fig. 1f) shows a wider $C_{\text{(NHC)}}\text{--Ru--S}$ angle of 148.0° (versus 143.4° for **Ru-3a**¹⁰); this indicates a lowering of *trans* influence between the NHC and the anionic ligand, presumably due to diminished electron density at the sulfide site.

Efficiency of the cross-metathesis to generate Z allylic alcohol **3** improved noticeably when the dichloro complex **Ru-3b** was used (77% versus 50% with **Ru-3a**; entry 1 of Table 1), and the desired product was obtained in 61% yield with minimal decrease in Z selectivity (96:4 versus 98:2 Z:E with **Ru-3a**). Dibromo (entry 2 of Table 1), chlorobromo (entry 3 of Table 1) and tetrafluoro (entry 4 of Table 1) complexes **Ru-3c** to **Ru-3e** were equally effective and stereoselectivities were high across the board (94:6–96:4 Z:E). However, **Ru-3b** has a practical advantage: unlike the other halogenated catechothiols, dichlorocatechothiol **4** (Fig. 1f) is commercially available. The lower efficiency of 2,5-dimethylcatechothiolate **Ru-3f** (entry 5 of Table 1; 56% conversion, 47% yield) is consistent with the positive electronic effect of the halogen substituents in **Ru-3b** to **Ru-3e**. The use of 5.0 mol% **Ru-3b**

a Comparison of the activity of **Ru-2** and **Ru-3b** complexes in Z-selective cross-metathesis



b Application to stereoselective synthesis of neopeltolide and leucascandrolide A

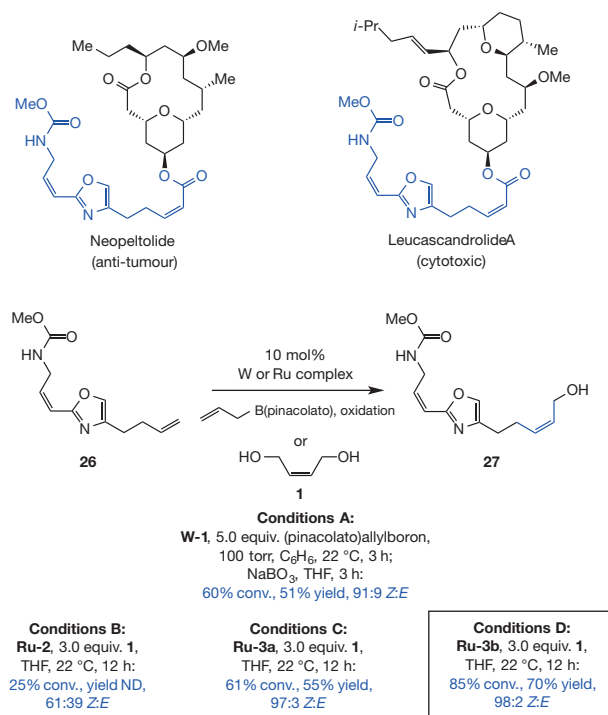


Figure 3 | Assessing the performance of different Z-selective catalysts.

a, A representative set of products that allow comparison of the two Ru-based carbene systems. Reactions with **Ru-2** can be more sensitive to the size of the substituent on the reaction partner (see **14** and **21**), and cross-metathesis with dienes can be more efficient with **Ru-3b** (see **22**). Unlike **Ru-3b**, when **Ru-2** is employed, the presence of an aldehyde (see **15**) lowers reaction rates and carboxylic acids (see **16**) completely shut down reactivity. All processes were performed (in duplicate or triplicate) with 5.0 mol% of the Ru complex at 22°C for four hours; see the Supplementary Information for details. **b**, A key step in the stereoselective synthesis of the side chains of the anti-tumour and cytotoxic natural products neopeltolide and leucascandrolide entails a site- and

Z-selective cross-metathesis between the terminal alkene in **26** and an appropriate cross partner. The optimal high-oxidation-state W-based alkylidene (10 mol%) requires the two-step process involving cross-metathesis with (pinacolato)allylboron followed by C–B bond oxidation to afford **27** in 51% yield and a 91:9 Z:E mixture of difficult-to-separate alkene isomers. With **Ru-2** and 3.0 equivalents of Z-2-butene-1,4-diol **1** there is only 25% conversion to the desired product, which is formed in low stereoselectivity (61:39 Z:E). Reaction with **Ru-3a** is highly Z-selective but proceeds to $\sim 60\%$ conversion. In contrast, when **Ru-3b** was used, Z allylic alcohol **27** is obtained in 70% yield and 98:2 Z:E ratio. ND, not determined. See the Supplementary Information and ref. 25 for details of conditions A to D.

led to 84% conversion within four hours, delivering **3** in 71% yield and 96:4 *Z:E* selectivity (entry 6 of Table 1).

The catalytic protocol exhibits considerable scope (Fig. 2). Reactions with various terminal alkenes, including those with heterocyclic and/or Lewis basic functional groups, afforded *Z* allylic alcohols **6–12** in 57–80% yield and 91:9–98:2 *Z:E* selectivity. *Z* allylic alcohol **6** has been used to prepare (+)-disparlure, a female sex pheromone of the gypsy moth *Porthetria dispar* (L.)²¹. To demonstrate practicality, the transformation leading to the formation of phthalamide **11** was successfully performed on the gram scale. The *Z*-selective cross-metathesis leading to diol **13** took place with useful efficiency and stereoselectivity levels (64% yield, 96:4 *Z:E*). The transformation that delivered **14** (68% yield, 98:2 *Z:E*) involves a substrate with a particularly acidic phenolic unit (owing to a *para* ketone). Synthesis of aldehyde-substituted *Z* allylic alcohol **15** was efficient (80% yield) and similarly selective (94:6 *Z:E*), as were the processes that generated carboxylic acid **16** (70% yield, 96:4 *Z:E*) or **17** (61% yield, 98:2 *Z:E*). The method is applicable to reactions with sizeable α -branched substrates (see **18–21**); alkyl- and aryl-substituted olefins participated in cross-metathesis reactions that afforded the desired products in 53–63% yield and 93:7–98:2 *Z:E* selectivity. Prolonged reaction times (that is, exceeding four hours) did not improve conversion. In the case of **20** and **21** slightly higher yields were obtained with **Ru-3d**. Stereochemically defined 1,3-*Z,E*-dienes **22–24** were generated with high stereoselectivity (92:8–96:4 *Z:E*); formation of the more sterically congested (*Z,Z*)-1,3-diene **25** proceeded in 87% *Z* selectivity. The only other examples of *Z*-selective cross-metathesis involving 1,3-dienes correspond to processes catalysed by Mo- or W-based alkylidenes with allyl- or alkenyl(pinacolato)boron compounds²², where pinacolato is (CH₃)₄C₂O₂.

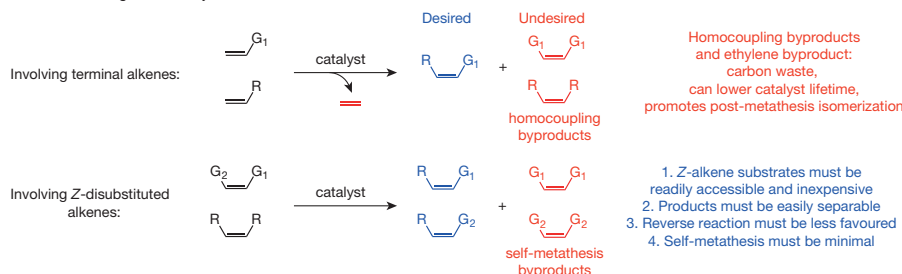
The above findings illustrate that replacing the chloride ligands of a Ru complex with an aryl disulfide unit does not alter the ability of the catalyst to promote cross-metathesis with sizeable substrates and/or those with a carbonyl or a Brønsted acid group. For comparison, we performed a number of the reactions with **Ru-2** (Fig. 3a). Allylic alcohol **14** was obtained in 50% yield and 82:18 *Z:E* selectivity when **Ru-2** was used (versus 68% yield and 98:2 *Z:E* with **Ru-3b**). There is a larger gap in efficiency with a sterically demanding aryl alkene, where 19% conversion

to **21** was observed after four hours in the case of **Ru-2** (55% yield and 93:7 *Z:E* with **Ru-3b**). Synthesis of 1,3-diene **22** proved to be somewhat more efficient with **Ru-3b** (66% versus 53% yield with **Ru-2**). Catalytic cross-metathesis affording aldehyde-containing allylic alcohol **15** proceeded in 30% yield and 87:13 *Z:E* selectivity with **Ru-2** (versus 86% conversion, 80% yield and 94:6 *Z:E* with **Ru-3b**). Carboxylic acids are detrimental to the catalytic activity of complexes represented by **Ru-2** (ref. 6): we were unable to observe any transformation in our attempts to prepare **16** or **17** (<2% conversion versus 70% and 61% yield, respectively, and \geq 96:4 *Z:E* with **Ru-3b**). It appears that the highly polarized and nucleophilic Ru–alkyl bond in **Ru-2** gives rise to complications with aldehyde- or carboxylic-acid-containing substrates, reminiscent of Mo- or W-based alkylidenes. All processes shown in Fig. 2 were performed under a N₂ atmosphere in a typical fume hood at ambient temperature (22 °C).

An occasion to challenge the utility of the cross-metathesis strategy presented itself in the course of the total synthesis of the anti-tumour agent neopeltolide²³; the same acyclic (*Z,Z*)-diene appendage is found in the cytotoxic natural product leucascandrolide A (ref. 24; Fig. 3b). We required the conversion of diene **26** to *Z*-allylic alcohol **27**; the optimal conditions involved a stereoselective cross-metathesis with (pinacolato)allylboron in the presence of 10 mol% **W-1** (ref. 22), followed by oxidation of the resulting *Z*-(pinacolato)allylboron product to deliver **27** in 51% overall yield and 91:9 *Z:E* selectivity. We sought a more efficient and selective process, particularly since the alkene isomers proved to be inseparable. With 10 mol% **Ru-2** and diol **1**, there was 25% conversion to **27** (12 h), which was formed with 61% *Z* selectivity. When catechothiolate **Ru-3a** was employed, **27** was isolated in 55% yield and 97% *Z* selectivity. In contrast, with **Ru-3b** there was 85% conversion to the desired product, which was isolated in 70% yield and 98:2 *Z:E* selectivity. Oxidation of the alcohol delivered the neopeltolide side chain, allowing for the completion of the total synthesis²⁵.

Another compelling area of application relates to transformations involving animal fats and vegetable oils, readily accessible and inexpensive compounds that represent viable substitutes for the dwindling petrochemicals²⁶. Such starting materials, most of which consist of carboxylic acids and alcohols (~80%), are being employed with increasing

a Different strategies for catalytic *Z*-selective cross-metathesis



b Stereoselective cross-metathesis with two readily accessible *Z*-alkenes to access higher-value *Z*-alkene products

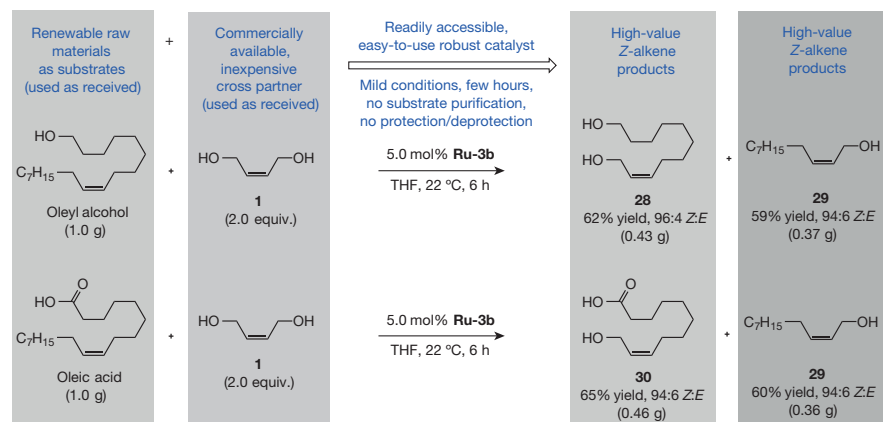


Figure 4 | Transformation of renewable raw materials to two different *Z* allylic alcohols by stereoselective cross-metathesis. **a**, Catalytic cross-metathesis of monosubstituted alkenes generates ethylene as a byproduct, which can lead to the formation of unstable methylidene complexes (reduced catalyst lifetime) with the ability to facilitate *Z*-to-*E* isomerization. In contrast, a catalyst may facilitate cross-metathesis between two *Z*-disubstituted olefins—one of which is structurally symmetric and both are easily accessible and inexpensive—to produce two new *Z* alkenes of higher value without homocoupling or detrimental effects of ethylene formation. **b**, Easily accessible and highly abundant plant oils (oleyl alcohol and oleic acid) can be converted to two high-value and readily separable *Z*-alkene products efficiently and stereoselectively. Transformations proceed readily at ambient temperature and pressure within six hours without the need for purification of the substrates or protection/deprotection of either the carboxylic acid or the hydroxyl units. Reactions with **Ru-3a** are less efficient and less stereoselective. G and R, functional groups.

regularity²⁷. Olefin metathesis catalysts that directly convert renewable natural resources to higher-value products efficiently and stereoselectively are expected to play a central role in the long-term future of diverse industries²⁷. Additionally, the combined use of two different 1,2-disubstituted alkene substrates offer definite advantages (over terminal alkenes; Fig. 4a). Yet, previous attempts at catalysing cross-metathesis with plentiful raw materials have led to inefficient and non-stereoselective transformations²⁸. Despite the use of dichloro–Ru complexes, reactions largely involved protected oleic acid derivatives. In a recent effort, cross-metathesis of oleic acid or methyl oleate with **1** or its bis-acetate derivative with a Ru-based complex at 50 °C afforded products as equilibrium mixtures of difficult-to-separate stereoisomers (70%–85% *E*)²⁸. Moreover, to ensure high activity, 100 equivalents (versus the Ru complex) of toxic PhSiCl₃ had to be present, and self-metathesis byproducts were formed (~20%)²⁸. In contrast, treatment of one gram of oleyl alcohol and **1** with 5.0 mol% **Ru-3b** for six hours at ambient temperature led to the formation of diol **28** (0.43 g, 62% yield) and allylic alcohol **29** (0.37 g, 59% yield); these easily separable products were formed in 96:4 and 94:6 *Z:E* selectivity, respectively (Fig. 4b). In the same fashion, a gram of oleic acid was converted to 0.46 g of acid-alcohol **30** (65% yield) with 94:6 *Z:E* selectivity and 0.36 g of **29** (94:6 *Z:E*). Carboxylic acid **30** is an anti-fungal agent²⁹, and allylic alcohol **29** is a substrate for stereoselective organic synthesis (see the Supplementary Information for further references).

Several additional points regarding the data in Fig. 4b are noteworthy. (1) Substrates were used as received from commercial vendors without purification; there was no need for rigorous removal of water or oxygen. (2) Cross-metathesis with the parent complex **Ru-3a** was markedly less efficient and stereoselective; for example, in the reaction with oleic acid, **30** and **29** were obtained in ~18–20% yield and ~86:14 *Z:E* ratio. (3) Side products from self-metathesis of oleyl alcohol or oleic acid were not detected. It is possible that formation of the hydroxymethylene-substituted carbene is strongly favoured (due in part to an internal H-bond¹¹); reaction of the latter species with **1** is degenerate and its transformation with the oleyl alcohol and oleic acid would result in the re-formation of the desired cross-metathesis products. However, such processes are no longer degenerate if they lead to generation of the *E* product isomer; the high *Z* selectivity even at advanced stages of the process underscores the fidelity of stereoselective cross-metathesis with Ru catechothiolate complexes.

The above attributes distinguish the present class of olefin metathesis catalysts not only from Mo- and W-based alkylidenes but from other types of Ru complexes as well. That is, the advantage of the Ru catechothiolate system extends beyond promoting olefin metathesis with stereochemical control, as the previous studies show that dichloro–Ru complexes (for example, **Ru-1**) cannot efficiently promote cross-metathesis of common disubstituted alkene feedstocks such as oleic acid (considerable substrate self-metathesis is observed)^{28,30}.

Received 15 September; accepted 4 November 2014.

- Hoveyda, A. H. & Zhugralin, A. R. The remarkable metal-catalyzed olefin metathesis reaction. *Nature* **450**, 243–251 (2007).
- Fürstner, A. Teaching metathesis “simple” stereochemistry. *Science* **341**, 1357–1364 (2013).
- Ibrahim, I., Yu, M., Schrock, R. R. & Hoveyda, A. H. Highly *Z*- and enantioselective ring-opening/cross-metathesis reactions catalyzed by stereogenic-at-Mo adamantylimido complexes. *J. Am. Chem. Soc.* **131**, 3844–3845 (2009).
- Meek, S. J., O'Brien, R. V., Lloveria, J., Schrock, R. R. & Hoveyda, A. H. Catalytic *Z*-selective olefin cross-metathesis for natural product synthesis. *Nature* **471**, 461–466 (2011).
- Yu, M. *et al.* Synthesis of macrocyclic natural products by catalyst-controlled stereoselective ring-closing metathesis. *Nature* **479**, 88–93 (2011).
- Keitz, B. K., Endo, K., Herbert, M. B. & Grubbs, R. H. *Z*-selective homodimerization of terminal olefins with a ruthenium metathesis catalyst. *J. Am. Chem. Soc.* **133**, 9686–9688 (2011).
- Keitz, B. K., Endo, K., Patel, P. R., Herbert, M. B. & Grubbs, R. H. Improved ruthenium catalysts for *Z*-selective olefin metathesis. *J. Am. Chem. Soc.* **134**, 693–699 (2012).
- Rosebrugh, L. E., Herbert, M. B., Marx, V. M., Keitz, B. K. & Grubbs, R. H. Highly active ruthenium metathesis catalysts exhibiting unprecedented activity and *Z* selectivity. *J. Am. Chem. Soc.* **135**, 1276–1279 (2013).

- Occhipinti, G., Hansen, F. R., Törnroos, K. W. & Jensen, V. R. Simple and highly *Z*-selective ruthenium-based olefin metathesis catalyst. *J. Am. Chem. Soc.* **135**, 3331–3334 (2013).
- Khan, R. K. M., Torker, S. & Hoveyda, A. H. Readily accessible and easily modifiable Ru-based catalysts for efficient and *Z*-selective ring-opening metathesis polymerization and ring-opening/cross-metathesis. *J. Am. Chem. Soc.* **135**, 10258–10261 (2013).
- Koh, M. J., Khan, R. K. M., Torker, S. & Hoveyda, A. H. Broadly applicable *Z*- and diastereoselective ring-opening/cross-metathesis catalyzed by a dithiolate Ru complex. *Angew. Chem. Int. Ed.* **53**, 1968–1972 (2014).
- Hoveyda, A. H. Evolution of catalytic stereoselective olefin metathesis. From ancillary transformation to purveyor of stereochemical identity. *J. Org. Chem.* **79**, 4763–4792 (2014).
- Mann, T. J., Speed, A. W. H., Schrock, R. R. & Hoveyda, A. H. Catalytic *Z*-selective cross-metathesis with secondary silyl- and benzyl-protected allylic ethers: mechanistic aspects and applications to natural product synthesis. *Angew. Chem. Int. Edn* **52**, 8395–8400 (2013).
- Buchmeiser, M. R., Sen, S., Unold, J. & Frey, W. N-Heterocyclic carbene, high oxidation state molybdenum alkylidene complexes: Functional-group-tolerant cationic metathesis catalysts. *Angew. Chem. Int. Edn* **53**, 9384–9388 (2014).
- Lin, Y. A. & Davis, B. G. The allylic chalcogen effect in olefin metathesis. *J. Org. Chem.* **6**, 1219–1228 (2010).
- Torker, S., Khan, R. K. M. & Hoveyda, A. H. The influence of anionic ligands on stereoisomerism of Ru carbenes and their importance to efficiency and selectivity of catalytic olefin metathesis reactions. *J. Am. Chem. Soc.* **136**, 3439–3455 (2014).
- Werner, H., Grünwald, C., Stürer, W. & Wolf, J. Deactivation of the Grubbs carbene complex [RuCl₂(=CHPh)(PCy₃)₂] by allylic alcohols. *Organometallics* **22**, 1558–1560 (2003).
- Hoveyda, A. H., Lombardi, P. J., O'Brien, R. V. & Zhugralin, A. R. H-bonding as a control element in stereoselective Ru-catalyzed olefin metathesis. *J. Am. Chem. Soc.* **131**, 8378–8379 (2009).
- Cannon, J. S. & Grubbs, R. H. Alkene chemoselectivity in ruthenium-catalyzed *Z*-selective olefin metathesis. *Angew. Chem. Int. Edn* **52**, 9001–9004 (2013).
- Hartung, J. & Grubbs, R. H. Catalytic, enantioselective synthesis of 1,2-anti-diols by asymmetric ring-opening/cross-metathesis. *Angew. Chem. Int. Ed.* **53**, 3885–3888 (2014).
- Rossiter, B. E., Katsuki, T. & Sharpless, K. B. Asymmetric epoxidation provides shortest routes to four chiral epoxy alcohols which are key intermediates in syntheses of methymycin, erythromycin, leukotriene C-1, and disparlure. *J. Am. Chem. Soc.* **103**, 464–465 (1981).
- Kiesewetter, E. T. *et al.* Synthesis of *Z*-(pinacolato)allylboron and *Z*-(pinacolato)alkenylboron compounds through stereoselective catalytic cross-metathesis. *J. Am. Chem. Soc.* **135**, 6026–6029 (2013).
- Wright, A. E. *et al.* Neopeltolide, a macrolide from Lithistid sponge of the family Neopeltidae. *J. Nat. Prod.* **70**, 412–416 (2007).
- D'Ambrosio, M., Guerriero, A., Debitus, C. & Pietra, F. Leucascandrolide A, a new type of macrolide: the first powerfully bioactive metabolite of calcareous sponges (*Leucascandra caveolata*, a new genus from the coral sea). *Helv. Chim. Acta* **79**, 51–60 (1996).
- Miao, Y., Schrock, R. R. & Hoveyda, A. H. Catalyst-controlled stereoselective olefin metathesis as a principal strategy in multi-step synthesis design. A concise route to (+)-neopeltolide. *Angew. Chem. Int. Ed.* <http://dx.doi.org/doi:10.1002/anie.201409120> (2014).
- Biermann, U., Bornscheuer, U., Meier, M. A. R., Metzger, J. & Schäfer, H. Oils and fats as renewable raw materials in chemistry. *Angew. Chem. Int. Ed.* **50**, 3854–3871 (2011).
- Gunstone, F. D. in *Oleochemical Manufacture and Applications* (eds Gunstone, F. D. & Hamilton, R. J.) Vol. 1 (Academic Press, 2001).
- Behr, A. & Gomes, J. P. The cross-metathesis of methyl oleate with *cis*-2-butene-1,4-diyl acetate and the influence of protecting groups. *J. Org. Chem.* **7**, 1–8 (2011).
- Suzuki, Y., Kurita, O., Kono, Y., Hyakutake, H. & Sakurai, A. Structure of a new antifungal C11-hydroxyfatty acid isolated from leaves of wild rice (*Oryza officinalis*). *Biosci. Biotechnol. Biochem.* **59**, 2049–2051 (1995).
- Kajetanowicz, A., Sytniczuk, A. & Grela, K. Metathesis of renewable raw materials— influence of ligands in the indenylidene type catalysts on self-metathesis of methyl oleate and cross-metathesis of methyl oleate with (*Z*)-2-butene-1,4-diol diacetate. *Green Chem.* **16**, 1579–1585 (2014).

Supplementary Information is available in the online version of the paper.

Acknowledgements This research was supported by a grant from the National Science Foundation (CHE-1362763). R.K.M.K. and M.Y. were partially supported as AstraZeneca Graduate Fellows. We thank Boston College for access to computational facilities.

Author Contributions M.J.K. and R.K.M.K. carried out the catalyst synthesis, method development studies and applications related to renewable feedstock, S.T. performed the computational investigations, M.Y. carried out the experiments in connection with neopeltolide, and M.S.M. studied modes of catalyst decomposition. A.H.H. conceived and directed the investigations and composed the manuscript with revisions provided by the other authors.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to A.H.H. (amir.hoveyda@bc.edu).

The geographical distribution of fossil fuels unused when limiting global warming to 2 °C

Christophe McGlade¹ & Paul Ekins¹

Policy makers have generally agreed that the average global temperature rise caused by greenhouse gas emissions should not exceed 2 °C above the average global temperature of pre-industrial times¹. It has been estimated that to have at least a 50 per cent chance of keeping warming below 2 °C throughout the twenty-first century, the cumulative carbon emissions between 2011 and 2050 need to be limited to around 1,100 gigatonnes of carbon dioxide (Gt CO₂)^{2,3}. However, the greenhouse gas emissions contained in present estimates of global fossil fuel reserves are around three times higher than this^{2,4}, and so the unabated use of all current fossil fuel reserves is incompatible with a warming limit of 2 °C. Here we use a single integrated assessment model that contains estimates of the quantities, locations and nature of the world's oil, gas and coal reserves and resources, and which is shown to be consistent with a wide variety of modelling approaches with different assumptions⁵, to explore the implications of this emissions limit for fossil fuel production in different regions. Our results suggest that, globally, a third of oil reserves, half of gas reserves and over 80 per cent of current coal reserves should remain unused from 2010 to 2050 in order to meet the target of 2 °C. We show that development of resources in the Arctic and any

increase in unconventional oil production are incommensurate with efforts to limit average global warming to 2 °C. Our results show that policy makers' instincts to exploit rapidly and completely their territorial fossil fuels are, in aggregate, inconsistent with their commitments to this temperature limit. Implementation of this policy commitment would also render unnecessary continued substantial expenditure on fossil fuel exploration, because any new discoveries could not lead to increased aggregate production.

Recent climate studies have demonstrated that average global temperature rises are closely related to cumulative emissions of greenhouse gases emitted over a given timeframe^{2,6,7}. This has resulted in the concept of the remaining global 'carbon budget' associated with the probability of successfully keeping the global temperature rise below a certain level^{4,8,9}. The Intergovernmental Panel on Climate Change (IPCC)³ recently suggested that to have a better-than-even chance of avoiding more than a 2 °C temperature rise, the carbon budget between 2011 and 2050 is around 870–1,240 Gt CO₂.

Such a carbon budget will have profound implications for the future utilization of oil, gas and coal. However, to understand the quantities that are required, and are not required, under different scenarios, we first

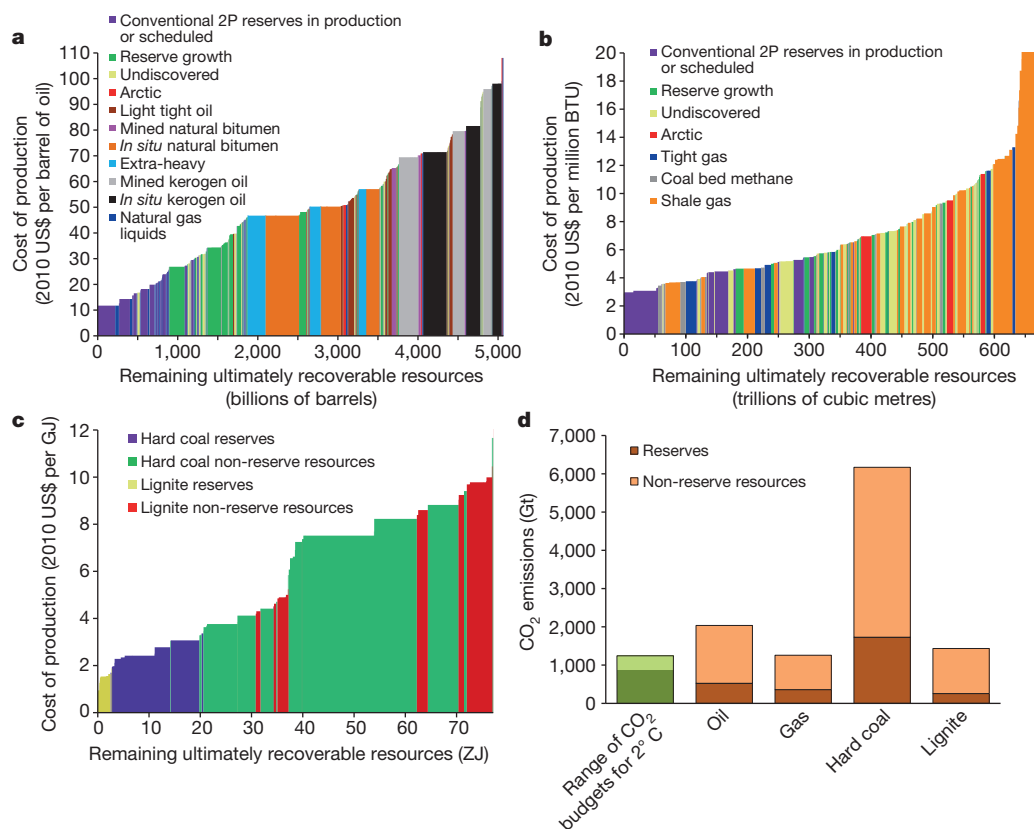


Figure 1 | Supply cost curves for oil, gas and coal and the combustion CO₂ emissions for these resources. a–c, Supply cost curves for oil (a), gas (b) and coal (c). d, The combustion CO₂ emissions for these resources. Within these resource estimates, 1,294 billion barrels of oil, 192 trillion cubic metres of gas, 728 Gt of hard coal, and 276 Gt of lignite are classified as reserves globally. These reserves would result in 2,900 Gt of CO₂ if combusted unabated. The range of carbon budgets between 2011 and 2050 that are approximately commensurate with limiting the temperature rise to 2 °C (870–1,240 Gt of CO₂) is also shown. 2P, 'proved plus probable' reserves; BTU, British thermal units (one BTU is equal to 1,055 J). One zettajoule (ZJ) is equal to one sextillion (10²¹) joules. Annual global primary energy production is approximately 0.5 ZJ.

¹University College London (UCL), Institute for Sustainable Resources, Central House, 14 Upper Woburn Place, London WC1H 0NN, UK.

need to establish the quantities and location of those currently estimated to exist. A variety of metrics with disparate nomenclature are relied upon to report the availability of fossil fuels^{10,11}, but the two most common are 'resources' and 'reserves'. In this work 'resources' are taken to be the remaining ultimately recoverable resources (RURR)—the quantity of oil, gas or coal remaining that is recoverable over all time with both current and future technology, irrespective of current economic conditions. 'Reserves' are a subset of resources that are defined to be recoverable under current economic conditions and have a specific probability of being produced¹¹. Our best estimates of the reserves and resources are presented in Fig. 1 and, at the regional level, in Extended Data Table 1.

Figure 1 also compares the above carbon budget with the CO₂ emissions that would result from the combustion of our estimate of remaining fossil fuel resources (nearly 11,000 Gt CO₂). With the combustion emissions of the remaining reserves alone totalling nearly 2,900 Gt CO₂, the disparity between what resources and reserves exist and what can be emitted while avoiding a temperature rise greater than the agreed 2 °C limit is therefore stark.

Although previous research¹² has examined the implications that emissions mitigation might have on the rents collected by fossil fuel resource owners, more pertinent to policy and industry are the quantities of fossil fuel that are not used before 2050 in scenarios that limit the average global surface temperature rise to 2 °C. Such geographically disaggregated estimates of 'unburnable' reserves and resources are provided here using the linear optimization, integrated assessment model TIAM-UCL¹³.

To provide context to the issue of unburnable fossil fuels and our results, it is useful to examine scenarios provided by other models that quantify separately the volumes of oil, gas and coal produced globally under a range of future emissions trajectories⁵. Cumulative production between 2010 and 2050 from these are presented in Fig. 2. Since they have very different future greenhouse gas emissions profiles, we have converted them to approximate temperature rise trajectories. These have been calculated using the climate model MAGICC¹⁴, which generates a probability distribution over temperature rise trajectories for a given emissions profile. We use the 60th percentile temperature trajectory (to correspond with assumptions within TIAM-UCL) and then group the scenarios by the final temperature rise in 2100: below 2 °C, between 2 °C and 3 °C, or exceeding 3 °C.

In this work we have constructed three core scenarios that are constrained to limit the average surface temperature rise in all time periods to 2 °C, to 3 °C, and to 5 °C. Cumulative production of each fossil fuel between 2010 and 2050 in each of these scenarios can be identified within each of the three temperature groupings in Fig. 2.

The global reserves of oil, gas and coal included in Fig. 1 total approximately 7.4 ZJ, 7.1 ZJ and 20 ZJ, respectively. With narrow inter-quartile ranges, relative to the level of reserves available, Fig. 2 shows good agreement on the levels of fossil fuels produced within the temperature groups, despite the range of modelling methodologies and assumptions included.

Since assumptions in modelling the energy system are subject to wide bands of uncertainty¹⁵, we further constructed a number of sensitivity scenarios using TIAM-UCL that remain within a 2 °C temperature rise. These span a broad range of assumptions on production costs, the availability of bio-energy, oil and gas, demand projections, and technology availability (one with no negative emissions technologies, and one with no carbon capture and storage (CCS)) (Extended Data Table 2). The availability of CCS has the largest effect on cumulative production levels (Extended Data Fig. 1); however, there is little variability in the total production of fossil fuels if the world is to have a good chance of staying within the agreed 2 °C limit.

Global production of oil, gas and coal over time in our main 2 °C scenario is given in Fig. 3. This separates production by category, that is, by the individual kinds of oil and gas that make up the global resource base, and compares total production with the projections from the 2 °C scenarios in Fig. 2. The results generated using TIAM-UCL are a product

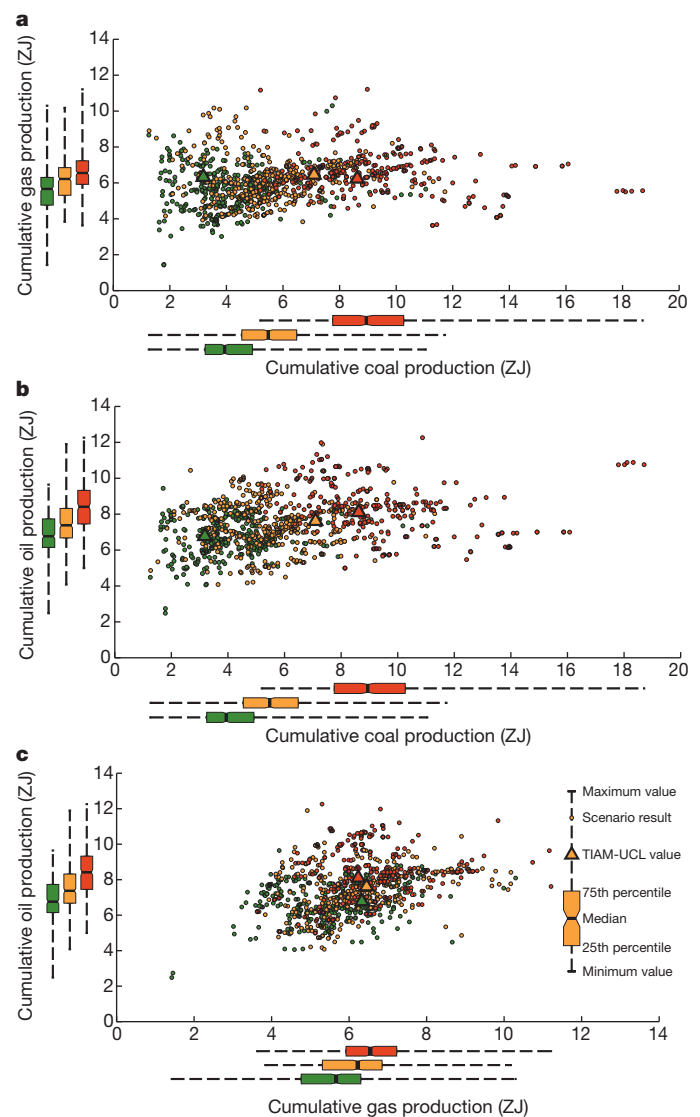


Figure 2 | Cumulative production between 2010 and 2050 from a range of long-term energy scenarios. Panels refer to coal and gas (a), coal and oil (b), and gas and oil (c). Scenarios⁵ are coloured according to their approximate resultant 2100 temperature rise above pre-industrial levels. 379 individual scenarios result in a temperature rise of less than 2 °C (green), 366 of between 2 °C and 3 °C (orange), and 284 of more than 3 °C (red). Triangles are the values from the 2 °C (with CCS), 3 °C and 5 °C TIAM-UCL scenarios. Ranges and symbols are as shown in the key in c.

of the economically-optimal solution, and other regional distributions of unburnable reserves are possible while still remaining within the 2 °C limit (even though these would have a lower social welfare). A future multi-model analysis could therefore usefully build on and extend the work that is presented here, but results at the aggregate level can be seen to lie within range of the ensemble of models and scenarios that also give no more than a 2 °C temperature rise.

In the TIAM-UCL scenarios, production of reserves and non-reserve resources occurs contemporaneously. It is therefore important to recognize that it would be inappropriate simply to compare the cumulative production figures in Fig. 2 with the reserve estimates from Fig. 1 and declare any reserves not used as 'unburnable'. Although there may be sufficient reserves to cover cumulative production between 2010 and 2050, it does not follow that only reserves should be developed and all other resources should remain unused. For oil and gas, resources that are not currently reserves may turn out to be cheaper to produce than some reserves, while new resources will also be developed to maintain

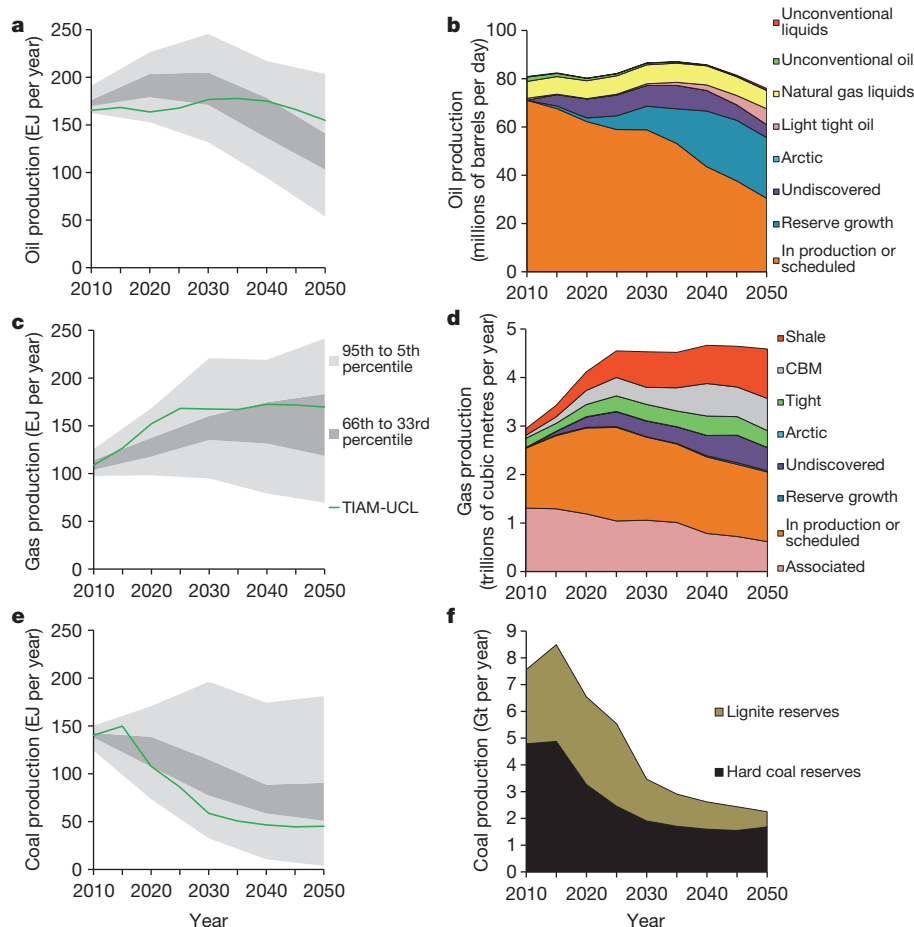


Figure 3 | Oil, gas and coal production in the TIAM-UCL 2 °C scenario (with CCS) and comparison with all other 2 °C scenarios in the Intergovernmental Panel on Climate Change Fifth Assessment Report (AR5) database⁵. a, c and e compare total production by oil, gas and coal with the AR5 database; b, d and f provide a disaggregated view of production for the TIAM-UCL 2 °C scenario separated by category. Associated gas is gas produced alongside crude oil from oil fields. One exajoule (EJ) is equal to one quintillion (10^{18}) joules.

the flow rates demanded by end-use sectors. However, if resources that are currently non-reserves are produced, a greater proportion of reserves must not be produced to stay within the carbon budget.

The reserves of oil, gas and coal that should be classified as unburnable within each region, and the percentage of current reserves that remain unused, are set out in Table 1. Since total production is most sensitive to assumptions on CCS, and since it has been suggested that the deployment of CCS will permit wider exploitation of the fossil fuel resource base¹⁶, Table 1 includes the unburnable reserves from two alternative 2 °C scenarios. One scenario permits the widespread deployment of CCS from 2025 onwards, and the other assumes that CCS is unavailable in any time period.

Globally, when CCS is permitted, over 430 billion barrels of oil and 95 trillion cubic metres of gas currently classified as reserves should remain

unburned by 2050. The Middle East, although using over 60% of its oil reserves, carries over half of the unburnable oil globally, leaving over 260 billions of barrels in the ground. Canada has the lowest utilization of its oil reserves (25%), as its natural bitumen¹⁷ deposits remain largely undeveloped (see below) while the United States has the highest, given the proximity of supply and demand centres. The Middle East also holds half of unburnable global gas reserves, with Former Soviet Union countries accounting for another third, meaning that they can use only half their current reserves.

Coal reserves are by far the least-used fossil fuel, with a global total of 82% remaining unburned before 2050. The United States and the Former Soviet Union countries each use less than 10% of their current reserves, meaning that they should leave over 200 billion tonnes (Gt) coal (both hard and lignite) reserves unburned. Coal reserve utilization

Table 1 | Regional distribution of reserves unburnable before 2050 for the 2 °C scenarios with and without CCS

Country or region	2 °C with CCS						2 °C without CCS					
	Oil		Gas		Coal		Oil		Gas		Coal	
	Billions of barrels	%	Trillions of cubic metres	%	Gt	%	Billions of barrels	%	Trillions of cubic metres	%	Gt	%
Africa	23	21%	4.4	33%	28	85%	28	26%	4.4	34%	30	90%
Canada	39	74%	0.3	24%	5.0	75%	40	75%	0.3	24%	5.4	82%
China and India	9	25%	2.9	63%	180	66%	9	25%	2.5	53%	207	77%
FSU	27	18%	31	50%	203	94%	28	19%	36	59%	209	97%
CSA	58	39%	4.8	53%	8	51%	63	42%	5.0	56%	11	73%
Europe	5.0	20%	0.6	11%	65	78%	5.3	21%	0.3	6%	74	89%
Middle East	263	38%	46	61%	3.4	99%	264	38%	47	61%	3.4	99%
OECD Pacific	2.1	37%	2.2	56%	83	93%	2.7	46%	2.0	51%	85	95%
ODA	2.0	9%	2.2	24%	10	34%	2.8	12%	2.1	22%	17	60%
United States of America	2.8	6%	0.3	4%	235	92%	4.6	9%	0.5	6%	245	95%
Global	431	33%	95	49%	819	82%	449	35%	100	52%	887	88%

FSU, the former Soviet Union countries; CSA, Central and South America; ODA, Other developing Asian countries; OECD, the Organisation for Economic Co-operation and Development. A barrel of oil is 0.159 m³; %, Reserves unburnable before 2050 as a percentage of current reserves.

is twenty-five percentage points higher in China and India, but still they should also leave nearly 200 Gt of their current coal reserves unburned.

The utilization of current reserves is lower in nearly all regions for all of the fossil fuels when CCS is not available, although there is a slight increase in gas production in some regions to offset some of the larger drop in coal production. Nevertheless, Table 1 demonstrates that the reserves of coal that can be burned are only six percentage points higher when CCS is allowed, with the utilization of gas and oil increasing by an even smaller fraction (around two percentage points). Because of the expense of CCS, its relatively late date of introduction (2025), and the assumed maximum rate at which it can be built, CCS has a relatively modest effect on the overall levels of fossil fuel that can be produced before 2050 in a 2 °C scenario.

As shown in Fig. 3, there is substantial production of many of the non-reserve resource categories of oil and gas. Extended Data Table 3 sets out the regional unburnable resources of all coal, gas and oil in the scenario that allows CCS by comparing cumulative production of all fossil fuel resources with the resource estimates in Fig. 1.

The RURR of both types of coal and unconventional oil vastly exceed cumulative production between 2010 and 2050, with the overwhelming majority remaining unburned. Resources of conventional oil are used to the greatest extent, with just under 350 billion barrels of non-reserve resources produced over the model timeframe. The Middle East again holds the largest share of the unburnable resources of conventional oil, but there is a much wider geographical distribution of these unburnable resources than was the case for oil reserves.

Regarding the production of unconventional oil, open-pit mining of natural bitumen in Canada soon drops to negligible levels after 2020 in all scenarios because it is considerably less economic than other methods of production. Production by *in situ* technologies continues in the 2 °C scenario that allows CCS, but this is accompanied by a rapid and total decarbonization of the auxiliary energy inputs required (Extended Data Fig. 2). Although such a decarbonization would be extremely challenging in reality, cumulative production of Canadian bitumen between 2010 and 2050 is still only 7.5 billion barrels. 85% of its 48 billion of barrels of bitumen reserves thus remain unburnable if the 2 °C limit is not to be exceeded. When CCS is not available, all bitumen production ceases by 2040. In both cases, the RURR of Canadian bitumen dwarfs cumulative production, so that around 99% of our estimate of its resources (640 billion barrels), remains unburnable. Similar results are seen for extra-heavy oil in Venezuela. Cumulative production is 3 billion barrels, meaning that almost 95% of its extra-heavy reserves and 99% of the RURR are unburnable, even when CCS is available.

The utilization of unconventional gas resources is considerably higher than unconventional oil. Under the 2 °C scenario, gas plays an important part in displacing coal from the electrical and industrial sectors and so there is over 50 trillion cubic metres unconventional gas production globally, over half of which occurs in North America. Nevertheless, there is a low level of utilization of the large potential unconventional gas resources held by China and India, Africa and the Middle East, and so over 80% of unconventional gas resources (247 trillion cubic metres) are unburnable before 2050. Production of these unconventional gas resources is, however, only possible if the levels of coal reserves identified in Table 1 are not developed: that is, it is not possible for unconventional gas to be additional to current levels of coal production.

Finally, we estimate there to be 100 billion barrels of oil (including natural gas liquids) and 35 trillion cubic metres of gas in fields within the Arctic Circle that are not being produced as of 2010. However, none is produced in any region in either of the 2 °C scenarios before 2050.

These results indicate to us that all Arctic resources should be classified as unburnable.

To conclude, these results demonstrate that a stark transformation in our understanding of fossil fuel availability is necessary. Although there have previously been fears over the scarcity of fossil fuels¹⁸, in a climate-constrained world this is no longer a relevant concern: large portions of the reserve base and an even greater proportion of the resource base should not be produced if the temperature rise is to remain below 2 °C.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 18 February; accepted 27 October 2014.

1. United Nations Framework Convention on Climate Change (UNFCCC). *Report of the Conference of the Parties on its Fifteenth Session, held in Copenhagen from 7 to 19 December 2009. Part Two: Action taken by the Conference of the Parties at its Fifteenth Session*. United Nations Climate Change Conf. Report 43 <http://unfccc.int/resource/docs/2009/cop15/eng/11a01.pdf> (UNFCCC, 2009).
2. Meinshausen, M. *et al.* Greenhouse gas emission targets for limiting global warming to 2 °C. *Nature* **458**, 1158–1162 (2009).
3. Clarke, L. *et al.* in *Climate Change 2014: Mitigation of Climate Change* (Edenhofer, O. *et al.*) Ch. 6 (Cambridge Univ. Press, 2014).
4. Raupach, M. R. *et al.* Sharing a quota on cumulative carbon emissions. *Nature Clim. Chang.* **4**, 873–879 (2014).
5. IPCC Working Group III. *Integrated Assessment Modelling Consortium (IAMC) AR5 Scenario Database* <https://secure.iiasa.ac.at/web-apps/ene/AR5DB/> (International Institute for Applied Systems Analysis, 2014).
6. Allen, M. R. *et al.* Warming caused by cumulative carbon emissions towards the trillionth tonne. *Nature* **458**, 1163–1166 (2009).
7. Matthews, H. D., Gillett, N. P., Stott, P. A. & Zickfeld, K. The proportionality of global warming to cumulative carbon emissions. *Nature* **459**, 829–832 (2009).
8. Friedlingstein, P. *et al.* Persistent growth of CO₂ emissions and implications for reaching climate targets. *Nature Geosci.* **7**, 709–715 (2014).
9. Leaton, J. *Unburnable Carbon—Are the World's Financial Markets Carrying a Carbon Bubble?* <http://www.carbontracker.org/wp-content/uploads/2014/09/Unburnable-Carbon-Full-rev2-1.pdf> (Investor Watch, 2011).
10. McGlade, C. E. A review of the uncertainties in estimates of global oil resources. *Energy* **47**, 262–270 (2012).
11. Society of Petroleum Engineers (SPE). *Petroleum Resources Management System*. www.spe.org/industry/docs/Petroleum_Resources_Management_System_2007.pdf (SPE, 2008).
12. Bauer, N. *et al.* Global fossil energy markets and climate change mitigation—an analysis with REMIND. *Clim. Change* <http://dx.doi.org/10.1007/s10584-013-0901-6> (2013).
13. Anandarajah, G., Pye, S., Usher, W., Kesicki, F. & McGlade, C. E. *TIAM-UCL Global Model Documentation*. <http://www.ucl.ac.uk/energy-models/models/tiam-ucl/tiam-ucl-manual> (University College London, 2011).
14. Meinshausen, M., Raper, S. C. B. & Wigley, T. M. L. Emulating atmosphere–ocean and carbon cycle models with a simpler model, MAGICC6—Part 1: Model description and calibration. *Atmos. Chem. Phys.* **11**, 1417–1456 (2011).
15. Usher, W. & Strachan, N. Critical mid-term uncertainties in long-term decarbonisation pathways. *Energy Policy* **41**, 433–444 (2012).
16. IEA. *Resources to Reserves* Ch. 8 (International Energy Agency, 2013).
17. Alberta Energy Regulator (AER). *Alberta's Energy Reserves 2013 and Supply/Demand Outlook 2014–2023*. www.aer.ca/documents/sts/ST98/ST98-2014.pdf (AER, 2014).
18. Yergin, D. *The Prize: the Epic Quest for Oil, Money and Power* Epilogue (Simon and Schuster, 2009).

Acknowledgements We thank I. Keppo at the UCL Energy Institute, E. Trutnevyte at ETH Zurich, and A.-M. Lyne at the UCL Department of Statistical Science. This research formed part of the programme of the UK Energy Research Centre and was supported by the UK Research Councils under Natural Environment Research Council award NE/G007748/1.

Author Contributions Both authors contributed equally to this work.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to C.M. (christophe.mcglade@ucl.ac.uk).

METHODS

Fossil fuel definitions. A 'McKelvey' box¹⁹ is often used to provide an overview of the relationship between different resource and reserve estimates²⁰. The best estimates of current oil and gas reserves in Extended Data Table 1 were of the 'proved plus probable' or '2P' quantities. Since 2P reserve estimates are rare for coal and none are in the public domain, the best estimates shown for coal were of the 'proved' or '1P' reserves. Broadly speaking, 1P estimates are more conservative, often corresponding to an estimate with a 90% probability of being exceeded, while 2P estimates are the median estimate of the reserves for a given field or region¹¹.

Oil and gas can be further separated into 'conventional' and 'unconventional' reserves and resources. Again, there is no single definition of these terms, but here we define oil with density greater than water (often standardized as '10° API') to be unconventional and all other quantities as conventional. We therefore categorize the 'light tight oil' extracted from impermeable shale formations using hydraulic fracturing as conventional oil.

For gas, tight gas (gas trapped in relatively impermeable hard rock, limestone or sandstone), coal-bed methane (gas trapped in coal seams that is adsorbed in the solid matrix of the coal), and shale gas (gas trapped in fine-grained shale) are considered as the three 'unconventional gases'; all other quantities are considered to be conventional.

Coal is distinguished by its energy density following the definitions used by the Federal Institute for Geosciences and Natural Resources (BGR)²¹. Hard coal has an energy density greater than 16.5 MJ kg⁻¹; any quantities with energy density less than this are classified as lignite.

Derivation of reserve and resource estimates. The estimated oil and gas reserves and resources shown in Extended Data Table 1 were derived in the following manner²². We first identified the individual elements or categories of oil and gas that make up the global resource base. For oil these are: current conventional 2P reserves in fields that are in production or are scheduled to be developed, reserve growth, undiscovered oil, Arctic oil, light tight oil, natural gas liquids, natural bitumen, extra-heavy oil, and kerogen oil. The latter three of these are the unconventional oil categories.

Reserve growth is defined to be 'the commonly observed increase in recoverable resources in previously discovered fields through time'²³. Quantities in this category here include any contributions from reserves in fields that have been discovered but are not scheduled to be developed ('fallow fields'), the new implementation of advanced production technologies such as enhanced oil recovery, changes in geological understanding, and changes in regional definitions.

There are eight categories of conventional and unconventional gas: current conventional 2P reserves that are in fields in production or are scheduled to be developed, reserve growth, undiscovered gas, Arctic gas, associated gas, tight gas, coal-bed methane, and shale gas. As noted above, the latter three of these are collectively referred to as unconventional gas.

We then selected the most robust data sources that provide estimates of the resource potential of each individual category within each country; these sources are set out in Extended Data Table 4. Taken together, differences between these sources provide a spread of discrete quantitative resource estimates for each category within each country. We also differentiated between the quantities of conventional oil that are natural gas liquids, and the quantities of natural gas that are associated with oil fields; these distinctions are important for modelling purposes but are rarely made in the literature.

For unconventional oil, we first generated a range of estimates for the in-place resources of natural bitumen, extra-heavy oil, and kerogen oil, and a range of potential recovery factors for different extraction technologies. We separately characterized the natural bitumen and kerogen oil resources that are extractable using mining technologies and those resources that are extractable using *in situ* technologies because the resource potential, costs, and energy requirements of these technologies are very different.

Continuous distributions were next constructed across these data ranges. Since there is no empirical basis for the choice of a suitable shape or form for such distributions, we used both the triangular and the beta distributions, chosen because they can be skewed both positively and negatively, and because they allow identical distributions to be used across all of the ranges derived. With equal weighting for each distribution, we combined these into a single individual resource distribution for each category within each country.

We then estimated the production costs of each of the oil and gas resource categories. Taking account of the resource uncertainty, these were used to develop supply cost curves for each category of oil and gas within each country.

We finally used a Monte Carlo selection process to combine these country-level supply cost curves. Regional supply cost curves were thus formed from aggregated supply cost curves for individual countries, and similarly supply cost curves formed for multiple categories of oil or gas within one or more countries. Data in Fig. 1 are the median values from these aggregate distributions with Extended Data Table 4

giving high (95th percentile), median, and low (5th percentile) estimates for each category at the global level.

In most industry databases of oil and gas reserves (for example, the database produced by the consultancy IHS CERA^{24,25}), some of the quantities classified as reserves lie in fields that were discovered over ten years ago, yet these fields have not been developed and there are no plans at present to do so. These are sometimes referred to as 'fallow fields'. For gas these quantities can also be called 'stranded gas', and they can be quite substantial; for example ref. 24 suggests that 50% gas reserves outside of North America are in stranded fields. Strictly, oil and gas in such fields should not be classified as reserves (for example, ref. 11 states that reserve quantities must have a 'reasonable timetable for development'). However, in this work, to ensure that the reserve estimates provided in Table 1 are not substantially different from the global totals provided by these industry databases, we follow their convention of classifying these quantities as reserves.

There are fewer independent estimates of reserves for coal and so we simply relied upon the estimates provided by the BGR²¹ for the reserve figures in Extended Data Table 1. The RRR of coal are more problematic to characterize, however. The 'resource' estimates provided by the BGR are not estimates of the quantities that can actually be extracted but are the in-place quantities; large portions of these are unlikely ever to be technically recoverable.

We therefore used the proved, probable and possible reserve estimates for hard coal and lignite provided by the World Energy Council²⁶ for a selection of countries. The sum of these three figures gives an estimate of the 'tonnage within the estimated additional amount in place that geological and engineering information indicates with reasonable certainty might be recovered in the future' (the definition provided by the World Energy Council). Since the sum of these three figures takes account of technical recoverability, we consider that, while imperfect, they provide a better estimate of the ultimately recoverable resources of coal than either the (narrower) proved reserve or the (broader) in-place resource estimates.

There are a number of countries that are estimated by the BGR to hold large quantities of coal in place but for which no probable and possible reserve estimates are provided by the World Energy Council. The ratio of the World Energy Council resource estimate to the BGR in-place estimate in countries that have estimates provided by both sources can vary substantially, but the average ratio is 16% for hard coal and 31% for lignite. We therefore assumed this ratio to generate resource estimates for all countries for which only BGR in-place estimates are provided. The proved reserve estimates of coal are so large themselves that the resource estimates are less important than is the case for oil and gas resource estimates.

There are few other sources providing a comprehensive overview of fossil fuel availability. Further, these often do not provide their sources or the methods used to generate estimates, do not define fully what categories or elements are included or excluded, and do not indicate sufficient conversion factors that would allow a like-with-like comparison. Some exceptions, however, are the IEA^{27,28}, the IIASA Global Energy Assessment (GEA)²⁹, and the BGR²¹. Their estimates are shown together with our aggregated reserve and resource estimates in Extended Data Table 5.

A number of factors contribute to the large variation between these estimates. A key reason is that the definitions of 'reserves' and 'resources' differ among sources, and so it is problematic to seek to compare them directly. For example, as noted above, the BGR, whose estimates are followed closely by the other sources, gives the total coal in place rather than an estimate of the resources that can be recovered, as in our study. Other reasons for the differences seen include: (1) the exclusion or inclusion of certain categories of fossil fuels such as light tight oil, aquifer gas, and methane hydrates; (2) whether proved (1P) or proved plus probable (2P) reserves are reported, and the methods used to generate the 1P reserve estimates; (3) the potential inflation of reserve estimates for political reasons, and whether they should consequently be increased or reduced³⁰; (4) the inclusion of stranded gas volumes in gas reserve estimates; (5) differences in the functional form used to estimate volumes of reserve growth (if reserve growth is included at all); (6) the difficulty in estimating current recovery factors (the ratio of recoverable resources to total resources in place), and how these may increase in the future; (7) differences between the methods used to estimate undiscovered oil and gas volumes; (8) the scarcity of reports providing reliable estimates of the potential resources of Arctic oil and gas, light tight oil, tight gas and coal bed methane, and the frequent consequent reliance upon expert judgement; (9) variation in what unconventional oil production technologies, which vary considerably in their recovery factors, will be used in the future; and (10) the chosen cut-off 'yield' (the volume of synthetic oil produced from a given weight of shale rock) for kerogen oil.

The estimates considered in our model are the result of careful and explicit consideration of all these issues, with our choices justified in the light of available knowledge. It can be seen in Extended Data Table 5, however, that our median figures are generally lower than the estimates provided by the other sources shown there. Therefore, although we consider our median resource estimates to be more robust than the figures used by these other sources, in fact these other estimates were found

to be closer to being correct, then the unburnable resources given in Extended Data Table 3 would also be larger. For example, if total gas resources are actually at the GEA high estimate, then the percentage that should be classified as unburnable before 2050 under the 2 °C scenario would increase to 99% rather than our estimate of 75%.

The cut-off date after which quantities that have not been produced should be considered 'unburnable' is also an important assumption. While there are no specific timeframes attached to the definition of reserves, quantities are usually required to be developed within, for example, a 'reasonable timeframe'¹¹. It is doubtful whether any reserves not produced by 2050 would fulfil this criterion. We therefore take cumulative production of reserves between 2010 and 2050 as the reserve 'utilization', and classify any quantities not used within this time as those that should be 'unburnable' if a certain temperature rise is not to be exceeded. Similarly, if none, or only a minor proportion, of a certain non-reserve resource is produced before 2050, then any current interest in developing it would be questionable. We thus also rely on 2050 as the cut-off date for classifying resources that should be considered as unburnable.

Description and key assumptions in TIAM-UCL. The TIMES Integrated Assessment Model in University College London ('TIAM-UCL') is a technology-rich, bottom-up, whole-system model that maximizes social welfare under a number of imposed constraints. It models all primary energy sources (oil, gas, coal, nuclear, biomass, and renewables) from resource production through to their conversion, infrastructure requirements, and finally to sectoral end-use. An extended explanation of input assumptions, approaches and data sources can be found in ref. 13. The base year of TIAM-UCL is 2005, the model is run in full to 2100, and thereafter the climate module is run to 2200. Results are presented here only between 2010 and 2050 (and are reported in five-year increments). All scenarios in this paper are run with the assumption of perfect foresight.

Resources and costs of all primary energy production are specified separately within 16 regions covering the world, and separately within the regions that contain members of the Organisation of Petroleum Exporting Countries (OPEC); the names of these are presented in Extended Data Table 6. For clarity in the main text, we have aggregated some of these regions into ten more-encompassing groups.

The climate module of TIAM-UCL is calibrated to the MAGICC model¹⁴. This module can be used to project the effects of greenhouse gas emissions on: atmospheric concentrations of greenhouse gas, radiative forcing, and average global temperature rises. It can also be used to constrain the model to certain bounds on these variables. In this work, the climate module is used to restrict the temperature rise to certain levels (as explained below). For the calibration to MAGICC, values from the probability distributions of climate parameters in MAGICC were selected so that there is a 60% chance that the temperature rise will remain below any level reported. Any constraints imposed using the TIAM-UCL climate module thus also correspond to this probability.

The emissions profiles⁵ used in Fig. 2 were converted to temperature rises using MAGICC. To ensure consistency with TIAM-UCL, we use the 60th percentile temperature trajectory from MAGICC and then group by the final temperature rise in 2100; there is therefore also a 60% chance that the temperature rise will be below the level indicated.

For each of the scenarios run in this paper using TIAM-UCL, a 'base case' is first formed that incorporates no greenhouse gas abatement policies. This base case uses the standard version of the model that relies upon minimizing the discounted system cost. This is used to generate base prices for each commodity in the model. TIAM-UCL is then re-run using the elastic-demand version with the greenhouse gas abatement policies introduced. This version of the model maximizes social welfare (the sum of consumer and producer surplus) and allows the energy-service demands to respond to changes in the endogenously determined prices resulting from these new constraints.

Fossil fuel modelling in TIAM-UCL. Oil and gas are both modelled in a similar manner in TIAM-UCL. The nine categories of conventional and unconventional oil and eight categories of conventional and unconventional gas identified above are all modelled separately. Coal production in TIAM-UCL is modelled more collectively, with only two categories, reserves and resources, for hard coal and lignite.

Natural bitumen and kerogen oil resources can be produced using either mining or *in situ* means, the technologies for which have different costs, efficiencies, and energy inputs. Although natural gas is predominantly used at present for the energy inputs to these unconventional resources, the model is free to choose any source of heat, electricity and hydrogen to allow greater flexibility. The costs of the auxiliary energy inputs required to extract and upgrade the native unconventional oils are determined endogenously by the model.

Each of the coal, gas and oil categories are modelled separately within the regions listed in Extended Data Table 6, with each resource category within each region split into three cost steps. As discussed above, the supply cost curves given in Fig. 1 comprise the data input to TIAM-UCL.

After processing, oil is next refined into products (gasoline, diesel, naphtha and so on), whereas processed gas and coal can be used directly. Fuel switching to and from all of the fossil fuels is possible. Trade of hard coal, crude oil, refined products, natural gas, both in pipelines and as liquefied natural gas, is allowed. Lignite cannot be traded between the regions.

Refined oil products can also be produced directly using Fischer–Tropsch processes with possible feedstocks of coal, gas, or biomass; these technologies can also be employed either with or without carbon capture and storage. Regional coal, oil and gas prices are generated endogenously within the model. These incorporate the marginal cost of production, scarcity rents, rents arising from other imposed constraints, and transportation costs.

A new key aspect of TIAM-UCL is the imposition of asymmetric constraints on the rate of production of oil and gas given a certain resource availability; these are intended to represent 'depletion rate constraints'. In TIAM-UCL, these constraints are modelled through introducing maximum annual production growth and maximum 'decline rate' restrictions. These are imposed on each cost step of each category of both oil and gas in each region, and ensure that the production follows a more realistic profile over time.

Data for these constraints are available at the field level from the bottom-up economic and geological oil field production model ('BUEGO')³¹. BUEGO contains a data-rich representation of 7,000 producing 'undiscovered' and discovered but undeveloped oil fields. These data include each field's 2P reserves, potential production capacity increases, water depth, capital and operating costs, and natural decline rate (the rate at which production would decline in the absence of any additional capital investment).

We used production-weighted averages (as of 2010) of the individual fields within each region to give average regional natural decline rates, which were imposed as maximum decline constraints in TIAM-UCL in the form of equal maximum annual percentage reductions. Although data on gas natural decline rates are much more sparse, some are available at a regional level³², which can be compared with similar results for oil natural decline rates²⁵. This comparison suggests that gas natural decline rates are on average 1% per year greater than for oil, with similar distributions for location (onshore/offshore) and size. The constraints placed on the maximum annual reductions in natural gas production were thus assumed to be 1% higher than those derived for oil.

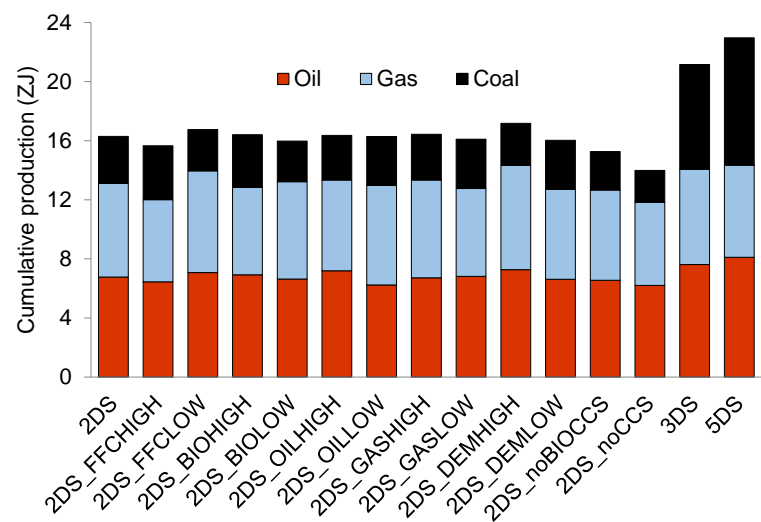
As identified in the main text, to understand the quantities of reserves of oil and gas that are unburnable, production of reserve sources only should be compared with reserve estimates, while cumulative production of all sources should be compared with the resource estimates. For coal, the reserves are so much greater than cumulative production under any scenario that this distinction is not as important.

The base year of TIAM-UCL is 2005, but the base year of this study is 2010. Since reserves have grown, and oil and gas have been discovered in the intervening five years, some quantities that were classified as reserve growth and undiscovered oil and gas in 2005 should be classified as reserves in 2010. Within each region, the cumulative production figures to which the reserve estimates in Extended Data Table 1 are compared therefore contain production from the conventional 2P reserves in the 'fields in production or scheduled to be developed' category, as well as some portions of production from the 'reserve growth' and 'undiscovered' categories. In addition, since, for example, reserves of natural bitumen are included in the reserves figures of Canada and unconventional gas reserves are included in the reserves figures of the United States, production of some of the unconventional categories are also included in these cumulative production figures. To ensure consistency within each region, the maximum production potentials over the modelling period from the categories included in the cumulative production figures are equal to the reserve estimates given in Extended Data Table 1.

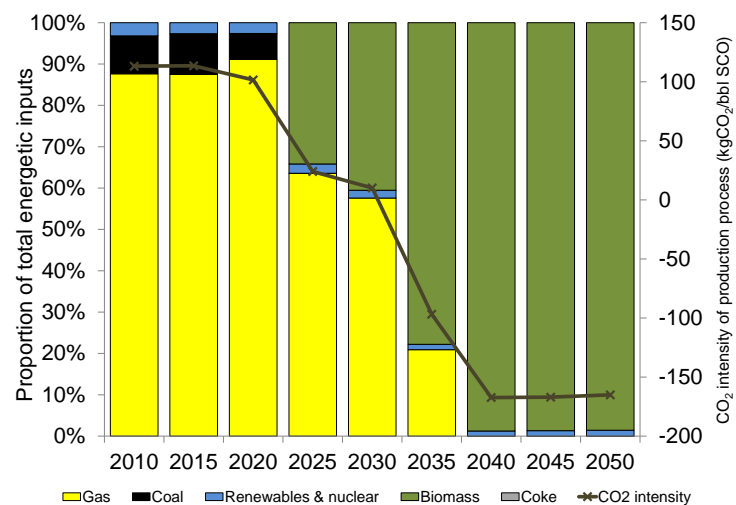
Overview of scenarios implemented. A brief overview of the main assumptions within the four scenarios run as part of this work is provided in Extended Data Table 7. For the emissions mitigation scenarios (those that limit the temperature rise to 3 °C and 2 °C), we assume that there are only relatively modest efforts to limit emissions in early periods as explained. The assumptions within the 2 °C sensitivity scenarios used to construct Extended Data Fig. 1 are provided in Extended Data Table 2.

19. McKelvey, V. E. Mineral resource estimates and public policy. *Am. Sci.* **60**, 32–40 (1972).
20. McGlade, C. E., Speirs, J. & Sorrell, S. Unconventional gas—a review of regional and global resource estimates. *Energy* **55**, 571–584 (2013).
21. Federal Institute for Geosciences and Natural Resources (BGR). *Energy Study 2012. Reserves, Resources and Availability of Energy Resources*. http://www.bgr.bund.de/DE/Gemeinsames/Produkte/Downloads/DERA_Rohstoffinformationen/rohstoffinformationen-15e.pdf?__blob=publicationFile&v=3 (BGR, 2012).
22. McGlade, C. E. *Uncertainties in the outlook for oil and gas*. PhD thesis, UCL, http://discovery.ucl.ac.uk/1418473/2/131106%20Christophe%20McGlade_PhD%20Thesis.pdf (2013).

23. Klett, T. & Schmoker, J. in *Giant Oil and Gas fields of the Decade 1990–1999* (ed. Halbouty, M. T.) 107–122 (The American Association of Petroleum Geologists, 2003).
24. Attanasi, E. D. & Freeman, P. A. Survey of Stranded Gas and Delivered Costs to Europe of Selected Gas Resources. *SPE Econ. Manag.* **3**, 149–162 (2011).
25. International Energy Agency (IEA). *World Energy Outlook*. <http://www.worldenergyoutlook.org/media/weowebsite/2008-1994/weo2008.pdf> (IEA, 2008).
26. Trinnaman, J. & Clarke, A. *Survey of Energy Resources* http://www.worldenergy.org/wp-content/uploads/2012/09/ser_2010_report_1.pdf (World Energy Council, 2010).
27. International Energy Agency (IEA). *World Energy Outlook*. <http://www.worldenergyoutlook.org/publications/weo-2013/> (IEA, 2013).
28. International Energy Agency (IEA). *World Energy Outlook*. http://www.iea.org/publications/freepublications/publication/weo2011_web.pdf (IEA, 2011).
29. Rogner, H.-H. *et al.* in *Global Energy Assessment—Towards a Sustainable Future* Ch. 7, 423–512 (Cambridge University Press, 2012).
30. Owen, N. A., Inderwildi, O. R. & King, D. A. The status of conventional world oil reserves—hype or cause for concern? *Energy Policy* **38**, 4743–4749 (2010).
31. McGlade, C. & Ekins, P. Un-burnable oil: an examination of oil resource utilisation in a decarbonised energy system. *Energy Policy* **64**, 102–112 (2014).
32. International Energy Agency (IEA). *World Energy Outlook*. <http://www.worldenergyoutlook.org/media/weowebsite/2009/WE02009.pdf> (IEA, 2009).
33. Leatherdale, A. *et al.* *Bioenergy Review: Technical Paper 2—Global and UK Bioenergy Supply Scenarios*. http://archive.theccc.org.uk/aws2/Bioenergy/1463%20CCC_Bio-TP2_supply-scen_FINALwithBkMks.pdf (Committee on Climate Change, 2011).
34. O'Neill, B. C. *et al.* A new scenario framework for climate change research: the concept of shared socioeconomic pathways. *Clim. Change* **122**, 387–400 (2014).
35. Campbell, C. J. *Atlas of Oil and Gas Depletion* (Springer, 2013).
36. Herrmann, L. *et al.* *Oil and Gas for Beginners* 270–413 (Deutsche Bank, 2013).
37. Klett, T. R. *et al.* An Assessment of Potential Additions to Conventional Oil and Gas Resources of the World (outside the United States) from Reserve Growth. <http://pubs.usgs.gov/fs/2012/3052/fs2012-3052.pdf> (USGS, 2012).
38. Klett, T. R. *et al.* *Potential Additions to Conventional Oil and Gas Resources in Discovered Fields of the United States from Reserve Growth, 2012*. <http://pubs.usgs.gov/fs/2012/3108/> (USGS, 2012).
39. Ahlbrandt, T., Charpentier, R., Klett, T., Schmoker, J. & Schenk, C. *USGS World Petroleum Assessment 2000*. <http://pubs.usgs.gov/dds/dds-060/> (USGS, 2000).
40. Bentley, R., Miller, R., Wheeler, S. & Boyle, G. *UKERC Review of Evidence on Global Oil Depletion: Annex 1—Models of global oil supply for the period 2008–2030*. http://www.ukerc.ac.uk/support/tiki-download_file.php?fileId=292 (UKERC, 2009).
41. Brownfield, M., Charpentier, R. R., Cook, T., Gautier, D. L. & Higley, D. K. *An Estimate of Undiscovered Conventional Oil and Gas Resources of the World, 2012*. <http://pubs.usgs.gov/fs/2012/3042/fs2012-3042.pdf> (USGS, 2012).
42. Gautier, D. L. *et al.* Assessment of undiscovered oil and gas in the Arctic. *Science* **324**, 1175–1179 (2009).
43. Smith, T. Arctic dreams—a reality check. *Geo ExPro* **4**, 16–24 (2007).
44. Shah, A. *et al.* A review of novel techniques for heavy oil and bitumen extraction and upgrading. *Energy Environ. Sci.* **3**, 700–714 (2010).
45. Clarke, B. *NPC Global Oil and Gas Study: Topic Paper 22—Heavy Oil*. www.npc.org/study_topic_papers/22-ttg-heavy-oil.pdf (National Petroleum Council, 2007).
46. Schenk, C. *et al.* *An Estimate of Recoverable Heavy Oil Resources of the Orinoco Oil Belt, Venezuela*. <http://pubs.usgs.gov/fs/2009/3028/pdf/FS09-3028.pdf> (USGS, 2009).
47. Attanasi, E. D. & Meyer, R. F. in *2010 Survey of Energy Resources* 123–150 (World Energy Council, 2010).
48. Johnson, R. C., Mercier, T. J. & Brownfield, M. Assessment of in-place oil shale resources of the Green River Formation, Greater Green River Basin in Wyoming, Colorado, and Utah. <http://pubs.usgs.gov/fs/2011/3063/pdf/FS11-3063.pdf> (USGS, 2011).
49. Dyni, J. *Geology and Resources of Some World Oil-Shale Deposits*. http://pubs.usgs.gov/sir/2005/5294/pdf/sir5294_508.pdf (USGS, 2006).
50. Biglarbigi, K., Mohan, H. & Carolus, M. *Potential for Oil Shale Development in the United States*. <http://www.inteki.com/reports.html> (INTEK, 2009).
51. CEDIGAZ. *Natural Gas in the World, End of July 2008* (Centre International d'Information sur le Gaz Naturel et tous Hydrocarbures Gazeux (CEDIGAZ), 2009).



Extended Data Figure 1 | Cumulative fossil fuel production under a range of sensitivity scenarios run using TIAM-UCL. Scenario names and characteristics are given in Extended Data Table 2.



Extended Data Figure 2 | The auxiliary energy inputs for natural bitumen production in Canada by *in situ* technologies in the 2 °C scenario and the CO₂ intensity of these. bbl SCO, a barrel of synthetic crude oil, the oil that results after upgrading the natural bitumen.

Extended Data Table 1 | Best estimates of remaining reserves and remaining ultimately recoverable resources from 2010

Country or region	Oil (Gb)			Gas (Tcm)			Hard coal (Gt)		Lignite (Gt)	
	Res	Con RURR	Uncon RURR	Res	Con RURR	Uncon RURR	Res	RURR	Res	RURR
Africa	111	280	70	13	45	35	31	45	2	5
Canada	53	60	640	1	5	25	4	35	2	40
China and India	38	90	110	5	10	40	255	1,080	16	120
FSU	152	370	360	61	95	30	123	580	94	490
CSA	148	360	450	9	30	55	10	25	5	10
Europe	25	110	30	6	25	20	17	70	66	160
Middle East	689	1,050	10	76	105	20	2	10	2	5
OECD Pacific	6	30	130	4	10	20	45	120	44	200
ODA	23	75	5	9	25	15	15	40	14	155
United States	50	190	650	8	25	40	226	560	31	335
Global	1,294	2,615	2,455	192	375	300	728	2,565	276	1,520

'Con' and 'Uncon' stand for conventional and unconventional sources, respectively. Coal is specified in billions of tonnes (Gt), gas in trillions of cubic metres (Tcm) and oil in billions of barrels (Gb). Res, reserves.

Extended Data Table 2 | Labels and description of the sensitivity scenarios modelled in this project

Sensitivity Name	Description
2DS_FFCHIGH	Production costs of all fossil fuel technologies are 50% larger in 2015 and 100% larger in 2020 than in 2DS, with equal annual percentage changes between these dates and remaining at this level for the model horizon
2DS_FFCLow	Production costs of all fossil fuel technologies are 33% lower in 2015 and 50% lower in 2020 than in 2DS, with equal annual percentage changes between these dates and remaining at this level for the model horizon
2DS_BIOHIGH	The maximum annual production of solid biomass and bio-crops in 2050 is assumed to be 350 EJ. This is close to the highest level of production of bio-energy in any of the scenarios from the AR5 scenario database ⁵ and is around three times the equivalent figure in 2DS (119 EJ).
2DS_BIOLOW	The maximum annual production of solid biomass and bio-crop in 2050 is assumed to be 38 EJ. This is similar to the figure given in the central scenario from ³³ and is around a third of the equivalent figure in 2DS (119 EJ).
2DS_OILHIGH	Uses the high values of each category of oil in each region from the aggregate resource distributions described in the methods section (Extended Data Table 4)
2DS_OILLOW	Uses the low values of each category of oil in each region (Extended Data Table 4)
2DS_GASHIGH	Uses the high values of each category of gas in each region (Extended Data Table 4)
2DS_GASLOW	Uses the low values of each category of gas in each region (Extended Data Table 4)
2DS_DEMHIGH	The major drivers of energy service demands in TIAM-UCL are growth in GDP, population, and GDP/capita. Future regional growth in GDP and population are therefore modified to the values given in Shared Socioeconomic Pathway (SSP) number 5 ³⁴ the SSP with the highest GDP and GDP/capita growth by 2050 (a 240% increase in the global average; cf. a 120% increase in 2DS). All other energy service demands (not relying on GDP or population) are also modified commensurately.
2DS_DEMLOW	Future regional growth in GDP and population are modified to the values given in Shared Socioeconomic Pathway (SSP) number 3: ³⁴ the SSP with the lowest GDP and GDP/capita growth by 2050 (a 50% increase in the global average).
2DS_NOBIOCCS	No negative emissions technologies are permitted i.e. carbon capture and storage (CCS) cannot be applied to any electrical or industrial process that uses biomass or bio-energy as feedstock in any period.
2DS_NOCCS	CCS is not permitted to be applied to any electrical or industrial process in any period.

Data for bio-energy sensitivities from refs 5 and 33, and for demand sensitivities from ref. 34.

Extended Data Table 3 | Regional distribution of resources unburnable before 2050 in absolute terms and as a percentage of current resources under the 2 °C scenario that allows CCS

Country or region	Conven oil		Unconven oil		Conven Gas		Unconven Gas		Hard Coal		Lignite	
	Gb	%	Gb	%	Tcm	%	Tcm	%	Gt	%	Gt	%
Africa	141	50%	70	100%	28	61%	35	100%	42	94%	2.8	56%
Canada	43	72%	633	99%	3.6	73%	18	71%	34	98%	39	97%
China and India	54	60%	110	100%	8.0	80%	35	88%	1,003	93%	106	88%
FSU	201	54%	360	100%	63	67%	27	89%	576	99%	480	98%
CSA	198	55%	447	99%	23	76%	51	92%	21	85%	6.3	63%
Europe	64	58%	30	100%	18	72%	16	78%	69	99%	142	89%
Middle East	554	53%	10	100%	72	68%	20	100%	10	100%	5.0	99%
OECD Pacific	23	77%	130	100%	9.0	90%	15	74%	116	97%	198	99%
ODA	38	51%	5.0	100%	14	55%	12	78%	34	84%	142	92%
United States	99	52%	650	100%	19	75%	20	50%	556	99%	317	95%
Global	1,417	54%	2,445	100%	257	69%	247	82%	2,462	96%	1,438	95%

*'Conven' and 'Unconven' stand for conventional and unconventional resources, respectively.

Extended Data Table 4 | Principal data sources used to derive reserve and resource estimates and estimates at the global level for each category of production

Category	Data sources used to provide country-level estimates of resources	Aggregated high estimate	Aggregated median estimate	Aggregated low estimate
Oil		(in Gb)	(in Gb)	(in Gb)
Current conventional 2P reserves in fields in production or scheduled to be developed	21,31,35,36	950	820	620
Reserve growth	37,38	1,200	850	610
Undiscovered oil	Fact sheets since USGS World Petroleum Assessment ³⁹ and ^{35,40,41}	580	300	180
Arctic oil	42,43	80	65	40
Light tight oil	10	470	300	150
Natural gas liquids (NGL)	26			
	Ancillary data associated with ³⁹	380	280	170
Natural bitumen	Oil in place estimates ^{17,26}	Mined RURR 130	Mined RURR 100	Mined RURR 70
	Extraction technologies ^{44–46}	<i>In situ</i> RURR 1290	<i>In situ</i> RURR 840	<i>In situ</i> RURR 520
Extra-heavy oil	Oil in place estimates ^{47,48}	750	440	230
	Extraction technologies ⁴⁷ and refs for bitumen			
Kerogen oil	Oil in place estimates ^{49,50}	Mined RURR 740	Mined RURR 485	Mined RURR 270
	Extraction technologies ⁵¹	<i>In situ</i> RURR 1,080	<i>In situ</i> RURR 590	<i>In situ</i> RURR 190
Total		7,650	5,070	3,050
Gas		(in tcm)	(in tcm)	(in tcm)
Current conventional 2P reserves in fields in production or scheduled to be developed	35,52	140	130	110
Reserve growth	24,37,38	125	90	60
Undiscovered gas	Fact sheets since USGS World Petroleum Assessment ³⁹ and ^{35,41}	180	120	80
Arctic gas	42,43	40	35	25
Tight gas	20	60	60	60
Coal-bed methane	20	45	40	20
Shale gas	20	310	200	120
Associated gas	36,37,44	Included in the above		
Total		900	675	475

High and low values are the aggregated 95th and 5th percentile estimates, respectively. 'tcm', trillions of cubic metres. Data are from references 10, 17, 20, 21, 31, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50 and 51.

Extended Data Table 5 | Global aggregated oil, gas and coal reserve and resource estimates from a selection of data sources

Organisation	Oil (Gb)		Gas (Tcm)		Coal (Gt)	
	Reserves	Resources	Reserves	Resources	Reserves	Resources
BGR	1,600	4,750	195	825	1,000	23,500
IEA	1,700	5,950	190	810	1,000	21,000
GEA	1,500 - 2,300	4,200 - 6,000	670 - 2,000	2,000 - 12,500	850 - 1,000	14,000 - 20,000
This study's median figures	1,300	5,070	190	675	1,000	4,085

BGR, Federal Institute for Geosciences and Natural Resources²¹; IEA, International Energy Agency^{27,28}; GEA, Global Energy Assessment²⁹.

Extended Data Table 6 | Regions included in TIAM-UCL and their aggregation to the regions given in the main text

Region	Aggregated region in main text
Non-OPEC Africa	Africa
OPEC Africa	Africa
Australia	OECD Pacific
Canada	Canada
Non-OPEC Central and South America	Central and South America (CSA)
OPEC Central and South America	Central and South America (CSA)
China	China and India
Eastern Europe	Europe
Former Soviet Union	Former Soviet Union (FSU)
India	China and India
Japan	OECD Pacific
Non-OPEC Middle	Middle East
OPEC Middle East	Middle East
Mexico	Central and South America (CSA)
Other Developing Asia	Other Developing Asia (ODA)
South Korea	OECD Pacific
United Kingdom	Europe
United States	United States
Western Europe	Europe

Extended Data Table 7 | Labels and description of the four core scenarios modelled in this project

Scenario Name	Description
5DS	<p>The model is constrained to keep the average global surface temperature rise to less than 5°C in all years to 2200.</p> <p>No other emissions constraints are imposed, and since allowed emissions under this scenario are so high (i.e. the constraint is very lax), no real emissions mitigation is required.</p> <p>These constraints result in 2050 GHG emissions of 71 Gt CO₂-eq (up from around 48 Gt CO₂-eq in 2010).</p>
3DS	<p>From 2005 to 2010, the model is fixed to the solution given in the 5°C temperature i.e. we assume that no emissions reductions are required.</p> <p>From 2010-2015, it is assumed that the model must be on track to achieve the emissions reduction pledges set out in the Copenhagen Accord¹, but no other emissions reductions are required.</p> <p>From 2015 onwards the model must meet the Copenhagen Accord emissions reductions in 2020, and emissions must be such as to keep the average global surface temperature rise below 3°C in all years to 2200.</p> <p>These constraints result in 2050 GHG emissions of 54 Gt CO₂-eq</p>
2DS	<p>The constraints between 2005 and 2015 in this scenario are identical to the 3DS.</p> <p>From 2015 onwards the model must meet the Copenhagen Accord emissions reductions in 2020, and emissions must be such as to keep the average global surface temperature rise below 2°C in all years to 2200.</p> <p>These constraints result in 2050 GHG emissions of 21 Gt CO₂-eq</p>
2DS-noCCS	<p>Emissions reduction requirements are identical to 2DS.</p> <p>Carbon capture and storage (CCS) is not permitted to be applied to any electricity or industrial process in any period.</p>

GHG, greenhouse gas measured in tonnes of CO₂ equivalent (CO₂-eq). Data from ref. 1.

Segmented lateral dyke growth in a rifting event at Bárðarbunga volcanic system, Iceland

Freysteinn Sigmundsson¹, Andrew Hooper², Sigrún Hreinsdóttir³, Kristín S. Vogfjörð⁴, Benedikt G. Ófeigsson⁴, Elías Rafn Heiðsson¹, Stéphanie Dumont¹, Michelle Parks¹, Karsten Spaans², Gunnar B. Gudmundsson⁴, Vincent Drouin¹, Thóra Árnadóttir¹, Kristín Jónsdóttir⁴, Magnús T. Gudmundsson¹, Thórdís Högnadóttir¹, Hildur María Fridriksdóttir^{1,4}, Martin Hensch⁴, Páll Einarsson¹, Eyjólfur Magnússon¹, Sergey Samsonov⁵, Bryndís Brandsdóttir¹, Robert S. White⁶, Thorbjörg Ágústsdóttir⁶, Tim Greenfield⁶, Robert G. Green⁶, Ásta Rut Hjartardóttir¹, Rikke Pedersen¹, Richard A. Bennett⁷, Halldór Geirsson⁸, Peter C. La Femina⁸, Helgi Björnsson¹, Finnur Pálsson¹, Erik Sturkell⁹, Christopher J. Bean¹⁰, Martin Möllhoff¹⁰, Aoife K. Braiden¹⁰ & Eva P. S. Eibl¹⁰

Crust at many divergent plate boundaries forms primarily by the injection of vertical sheet-like dykes, some tens of kilometres long¹. Previous models of rifting events indicate either lateral dyke growth away from a feeding source, with propagation rates decreasing as the dyke lengthens^{2–4}, or magma flowing vertically into dykes from an underlying source^{5,6}, with the role of topography on the evolution of lateral dykes not clear. Here we show how a recent segmented dyke intrusion in the Bárðarbunga volcanic system grew laterally for more than 45 kilometres at a variable rate, with topography influencing the direction of propagation. Barriers at the ends of each segment were overcome by the build-up of pressure in the dyke end; then a new segment formed and dyke lengthening temporarily peaked. The dyke evolution, which occurred primarily over 14 days, was revealed by propagating seismicity, ground deformation mapped by Global Positioning System (GPS), interferometric analysis of satellite radar images (InSAR), and graben formation. The strike of the dyke segments varies from an initially radial direction away from the Bárðarbunga caldera, towards alignment with that expected from regional stress at the distal end. A model minimizing the combined strain and gravitational potential energy explains the propagation path. Dyke opening and seismicity focused at the most distal segment at any given time, and were simultaneous with magma source deflation and slow collapse at the Bárðarbunga caldera, accompanied by a series of magnitude $M > 5$ earthquakes. Dyke growth was slowed down by an effusive fissure eruption near the end of the dyke. Lateral dyke growth with segment barrier breaking by pressure build-up in the dyke distal end explains how focused upwelling of magma under central volcanoes is effectively redistributed over long distances to create new upper crust at divergent plate boundaries.

The formation of dykes is favourable at divergent plate boundaries, because plate movements stretch the crust and reduce the normal stress on potential dyke planes. Rifting events at divergent plate boundaries typically occur in episodes separated by hundreds of years of quiescence. Only a few such episodes have been monitored, as most divergent plate boundaries form mid-ocean ridges. In 1975–84 a rifting episode took place at Krafla volcanic system, Iceland, and from 2005 to 2010 in the Afar region of Ethiopia¹. Limited geodetic and seismic data have been interpreted in terms of lateral flow of magma, with dyke propagation rates of up to 2–3 km per hour initially and then at a declining rate as magma propagates away from a central feeding source^{2–4}. The propagation of such dykes has been modelled as inflation of magma filled cracks with uniform excess pressure^{7,8}. The formation of regional dykes

in Iceland has alternatively been attributed to the vertical rise of magma from major magma reservoirs underlying dyke swarms^{5,6}.

Bárðarbunga is a subglacial basaltic central volcano with a 70 km² caldera at the northwestern corner of Vatnajökull ice cap in Iceland^{9,10} (Fig. 1, Extended Data Fig. 1). It has an associated fissure swarm¹¹ extending 115 km to the southwest and 55 km to the north-northeast. Activity in the last 2,000 years includes both subglacial eruptions and major effusive fissure eruptions, with 23 verified eruptions in the last 1,100 years¹². Timings of the most recent effusive eruptions north of the Vatnajökull ice cap, originating from the Bárðarbunga system, are not well known, but they are inferred to have produced the Holuhraun lava field sometime in the period from AD 1794 to 1864⁶. The Holuhraun eruptive fissure was reactivated in 2014. In 1996, the Gjalp subglacial eruption was likely to have been triggered by the Bárðarbunga volcanic system^{13,14}. Seismic activity at Bárðarbunga has been steadily increasing since 2005, mostly confined to the area northeast of its caldera.

On 16 August 2014 at 03:00 UTC, an intense seismic swarm began at Bárðarbunga. Initial seismic activity occurred in several clusters. One cluster was consistent with the formation of a radial dyke segment aligned in direction N127°E, outward from the Bárðarbunga caldera. Other clusters to the northeast of the caldera may also signify magma movements, or stress induced seismicity. GPS observations show simultaneous deflation of the caldera and displacements consistent with widening across the N127°E radial dyke, although deformation due to magma movements in the other clusters may also contribute. The seismic activity then focused on a lineament in direction N55°E, extending from the southern tip of the initial N127°E dyke segment (Extended Data Fig. 2). Lateral growth of this dyke is reflected in the migration of seismicity, along segments of variable strike; maximum widening of 1.3 m occurred between stations URHC and KVER spaced 25 km apart (Supplementary Fig. 1). Displacements of continuous GPS stations indicate the fastest rate of widening at any time in the most distal segment of the dyke throughout its evolution. The rate of dyke propagation varied considerably. A long halt in propagation, for 80 h, began on 19 August. Propagation rate exceeded 1 km h⁻¹ on 23 August when a new segment initiated with a 90° left turn and advanced 4 km north-northwest over two short segments. Following this the dyke took a right turn onto a new lineament striking N47°E, and then onto a N25°E striking segment.

The lengthening of the dyke ended on 27 August, around 10 km north of Vatnajökull, and a minor fissure erupted in Holuhraun for about 4 h on 29 August. On 31 August, a new eruption began from the same fissure and is still ongoing at the time of writing. After 4 September the

¹Nordic Volcanological Center, Institute of Earth Sciences, University of Iceland, IS-101 Reykjavík, Iceland. ²Centre for the Observation and Modelling of Earthquakes and Tectonics (COMET), School of Earth and Environment, University of Leeds, Leeds LS2 9JT, UK. ³GNS Science, PO Box 30368, Lower Hutt 5040, New Zealand. ⁴Icelandic Meteorological Office, IS-150 Reykjavík, Iceland. ⁵Canada Centre for Mapping and Earth Observation, Natural Resources Canada, 560 Rochester Street, Ottawa, Ontario K1A 0E4, Canada. ⁶Department of Earth Sciences, University of Cambridge, Madingley Road, Cambridge CB3 0EZ, UK. ⁷Department of Geosciences, University of Arizona, Tucson, Arizona 85721, USA. ⁸Department of Geosciences, The Pennsylvania State University, University Park, Pennsylvania 16802, USA. ⁹Department of Earth Sciences, University of Gothenburg, SE-405 30 Gothenburg, Sweden. ¹⁰Seismology Laboratory, School of Geological Sciences, University College Dublin, Belfield, Dublin 4, Ireland.

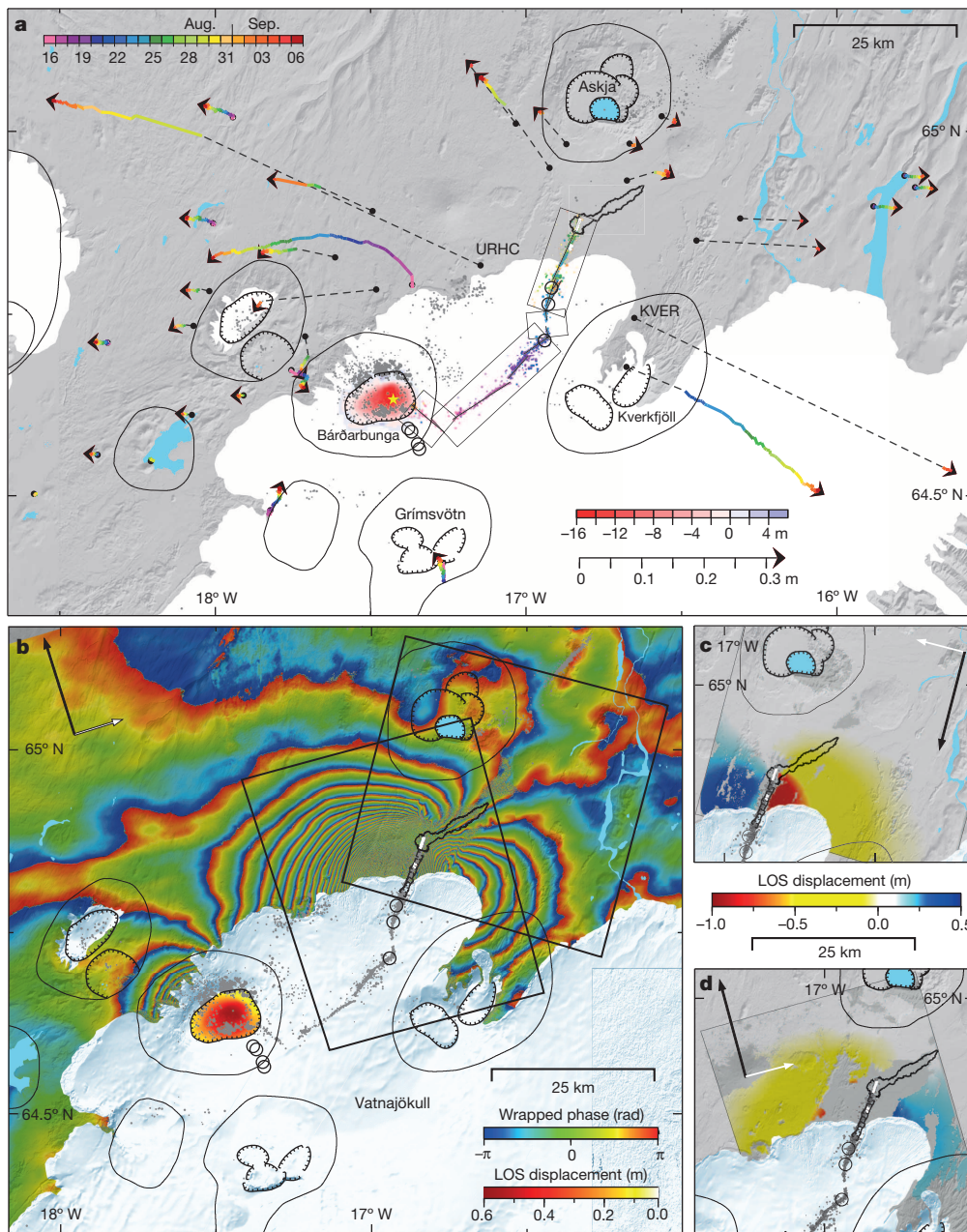


Figure 1 | Overview of data.
a, Earthquakes 16 August to 6 September 2014 (dots) and horizontal ground displacements measured by GPS (arrows) on a map with central volcanoes (oval outlines), calderas (hatched), and northern Vatnajökull. Relatively relocated epicentres and displacements are colour coded according to time of occurrence (key at top left), other single earthquake locations are in grey. Rectangles show areas displayed in Fig. 2; thin lines within them show inferred dyke segments. The red shading at Bárðarbunga caldera shows subsidence up to 16 m inferred from radar profiling on 5 September. The star marks the location of the magma source inferred from modelling. Also shown are ice cauldrons formed (circles), outline of lava flow mapped from radar image on 6 September, and eruptive fissures (white). **b**, Wrapped RADARSAT-2 interferogram spanning 8 August to 1 September 2014. Shading at Bárðarbunga caldera shows unwrapped one day (27–28 August) COSMO-SkyMed interferogram with maximum LOS (line of sight) increase of 57 cm. Also shown are earthquakes (grey dots), boundaries of graben activated in the dyke distal area (hatched lines), and location of interferograms in **c** and **d** (boxes). **c**, Unwrapped Cosmo-SkyMed interferogram spanning 13–29 August. **d**, Unwrapped TerraSAR-X interferogram spanning 26 July 2012 to 4 September 2014. Satellite flight and viewing direction are shown with black and white arrows, respectively. LOS displacement is positive away from the satellite for all interferograms shown.

movement associated with the dyke was minor, suggesting an approximate equilibrium between inflow of magma into the dyke and magma flowing out of it feeding the eruption. Minor eruptions may have occurred under Vatnajökull; shallow ice depressions marked by circular crevasses (ice cauldrons) were discovered in the period 27 August to 7 September, indicating leakage of magma or magmatic heat to the glacier causing basal melting (Figs 1 and 2b). On 5 September, aircraft radar profiling showed that the ice surface in the centre of the Bárðarbunga caldera had subsided 16 m relative to the surroundings, resulting in a $0.32 \pm 0.08 \text{ km}^3$ subsidence bowl (Fig. 1, Extended Data Fig. 3). No evidence for basal melting was observed inside the caldera, suggesting subsidence of the caldera floor. This slow collapse of the caldera floor is considered to have started between 16 August (start of unrest) and 24 August (beginning of a series of $M \geq 5$ earthquakes in the caldera), with an average rate of subsidence in this period of up to $0.8\text{--}1.2 \text{ m d}^{-1}$. It can be compared to a 1 day interferogram over the ice surface spanning 27–28 August (Fig. 1), which has a maximum line-of-sight (LOS) increase of 57 cm, indicating 55–70 cm of subsidence, over 24 h. From 24 August to 6 September, 16 $M \geq 5$ earthquakes occurred on the caldera boundary.

More than 22,000 earthquakes were automatically detected between 16 August and 6 September 2014, 5,000 of which have been manually checked. Four thousand of these have been relatively relocated, defining the dyke segments. Ground deformation in areas outside the Vatnajökull ice cap and on nunataks within the ice cap, is well mapped by a combination of InSAR, continuously recording GPS sites, and campaign GPS measurements. The GPS observations and analysis give the temporal evolution of the three-dimensional displacements used in the modelling (Fig. 1). Interferometric analysis of synthetic aperture radar images from the COSMO-SkyMed, RADARSAT-2, and TerraSAR-X satellites was used to form 11 interferograms showing LOS change spanning different time intervals (Supplementary Fig. 2). The analyses of seismic and geodetic data are described in Methods.

Initial modelling of the dyke, with no a priori constraints on position, strike or dip, shows that the deformation data require the dyke to be approximately vertical and line up with the seismicity (Extended Data Fig. 4). We therefore fixed the dip to be vertical and the lateral position of the dyke to coincide with the earthquake locations. We modelled the dyke as a series of rectangular patches and estimated the opening and

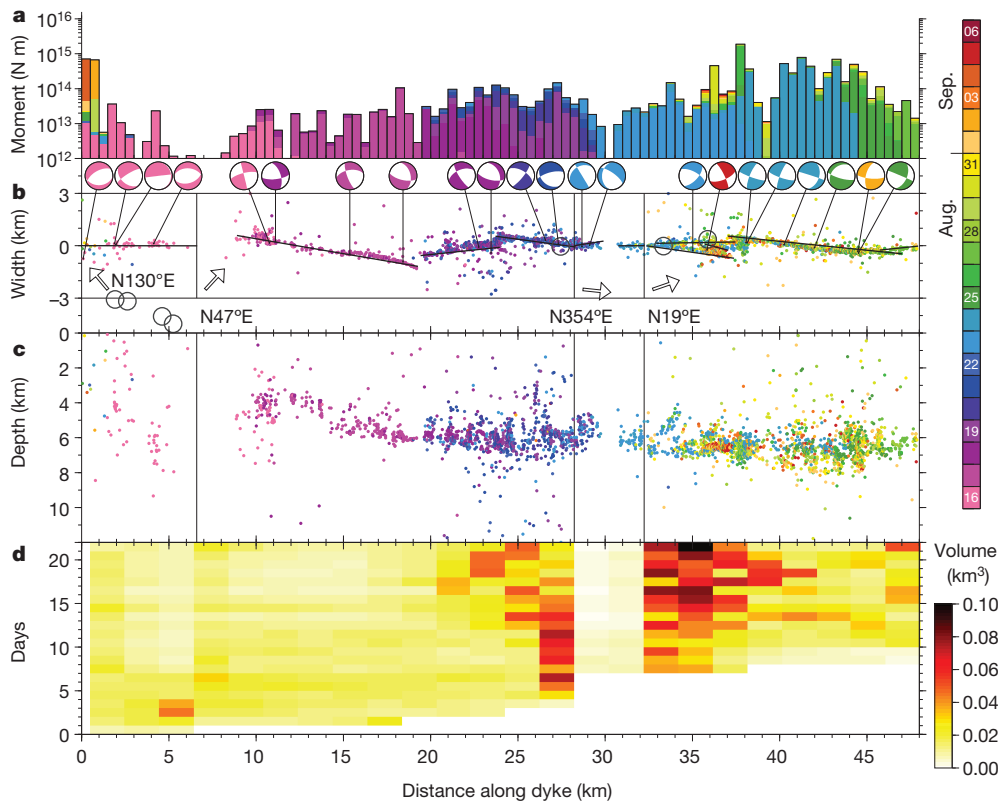


Figure 2 | Seismicity and magma volume along the dyke, 16 August to 6 September 2014. Relocated earthquakes shown in Fig. 1 are indicated, with same colour coding. **a**, Daily cumulative seismic moment at 0.5 km intervals along the dyke. **b**, Plan-view of four rotated areas along the dyke. Arrows indicate geographic north. Dots denote epicentres, black lines dyke segments, and open circles ice cauldrons. Fault-plane solutions for selected earthquakes are shown. **c**, Earthquake depths referenced to sea level. **d**, Daily vertically integrated volume of magma along the dyke inferred from geodetic modelling.

slip on each patch (Fig. 3a; see Supplementary Figs 3 and 4 for slip and standard deviations of opening). We used a Markov-chain Monte Carlo approach to estimate the multivariate probability distribution for all model parameters (Methods) on each day from 16 August to 6 September 2014 (Fig. 2d). The results suggest that most of the magma injected into the dyke is shallower than the seismicity, which mostly spans the depth range from 5 to 8 km below sea level (see Fig. 2c and Methods). While magma may extend to depths greater than 9 km near the centre of Bárðarbunga, towards the edge of the ice cap—where constraints from InSAR and GPS are much better—significant opening is all shallower than 5 km (Fig. 3a). The total volume intruded into the dyke by 28 August was 0.48–0.51 km³.

We took two approaches to deflation models: (1) by combining GPS displacements on 4 September, interferograms ending on 3 and 4 September, respectively, and the caldera subsidence measured on 5 September; and (2) by combining all data except the caldera subsidence in a time dependent model. Our approximate model has two dip-slip faults at the boundary of the caldera and an underlying magma source; either a spherical or a flat top chamber. In approach (1), the best-fit models have a spherical chamber centred at 1.3–1.5 km depth below sea level and a volume change of 0.26–0.29 km³, or a flat-topped chamber stretching from 3.4–3.6 km downwards and a volume change of 0.24–0.31 km³ (Extended Data Figs 5 and 6, Supplementary Figs 5 and 6). We consider the actual volume loss at depth to be at least equal to the volume of the caldera subsidence on 5 September (0.32 ± 0.08 km³); the volumes predicted by our simple models are marginally smaller. The time dependent models not using the caldera subsidence result in under-prediction of the volume change. Inverting the GPS and InSAR data from 3 and 4 September, but neglecting the caldera subsidence measurements, results in a volume change that is smaller by a factor of 2.0. We therefore scale the estimated volumes in our time dependent models by this factor to give more reliable estimate (Fig. 4). The volume decrease beneath the caldera tracks the volume increase of the dyke for the first week of the activity. The volume decrease then decelerates to less than half the previous rate, although the dyke volume increase continues at the same rate. This suggests inflow of magma from an underlying

deeper source after the first week, which is not visible in the geodetic data. Full details of the results are given in Methods and Supplementary Information.

Lateral growth of a dyke is expected to follow the minimum potential energy principle. Assuming a closed system, a dyke will tend to be emplaced such that it minimizes the total potential energy^{15–17}, Θ_T , given by:

$$\Theta_T = \Theta_s + \Theta_g \quad (1)$$

where Θ_s is the strain potential and Θ_g the gravitational energy potential. Evaluation of the strain energy requires knowledge of the prior stress and strain field in the crust. We consider here the role of plate movements and topography in steering the propagation path of a dyke once it is initiated; its onset point will depend on other factors, such as details of the magma plumbing system feeding it and the path of previous dykes. We approximate strain and stress due to plate movements as described in Methods, and then consider strain changes induced by the dyke formation. Opening of a dyke is energetically favourable when it releases strain energy built up at a divergent plate boundary, but once deviatoric stress in the crust adjacent to a segment is released it becomes favourable to propagate laterally. We estimate the total strain energy before and after advance of a dyke segment by numerically integrating the strain energy density over a large volume:

$$\Theta_s = \frac{1}{2} \int \sigma_{ij} \epsilon_{ij} dV \quad (2)$$

where σ_{ij} , ϵ_{ij} and dV are respectively the components of the stress tensor, strain tensor, and the volume element of integration¹⁵. We approximate the change in gravitational energy in surrounding crust, for each dyke segment, by integrating the predicted vertical displacements multiplied by the local topographic load density (ice and crust) above a reference surface and the acceleration of gravity (Methods). Dyke formation is associated with uplift on their flanks; the lower the topographic load over the flanks, the less energy it costs. For any given location on a volcano,

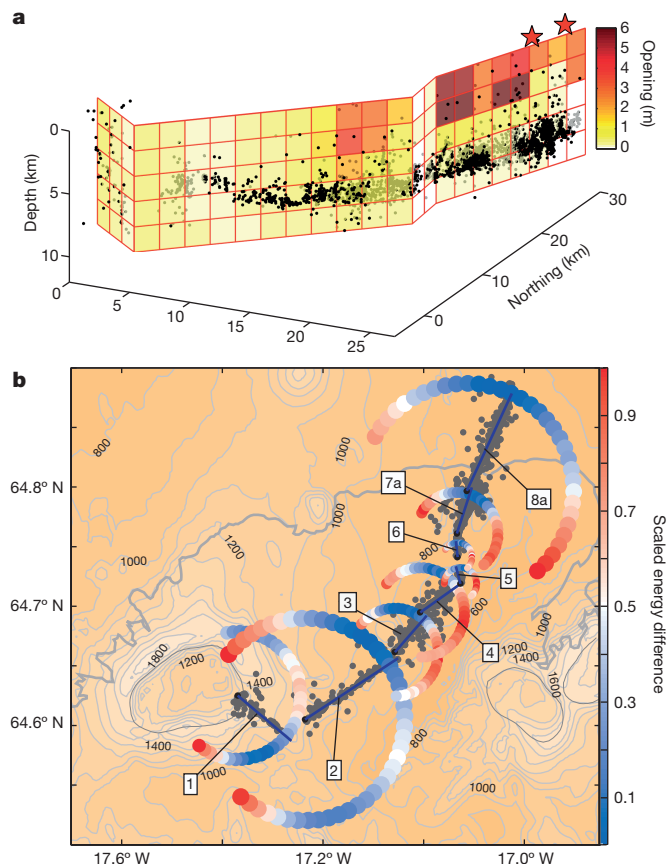


Figure 3 | Dyke model. **a**, Median of the posterior probability of opening for dyke patches inferred from modelling (key at right), and relocated earthquake hypocentres (black and grey dots, in front of and behind the dyke plane, respectively) relative to sea level. Red stars mark the eruption sites. **b**, Preferred direction of dyking for different segments based on a model of combined strain and gravitational potential energy release. Blue lines represent dyke segments and grey dots earthquake epicentres. Black dots indicate the beginning of each segment and surrounding arc of coloured points represent possible end points for different strikes of propagation. Their colour (key at right) indicates $(E - E_{\min}) / (E_{\max} - E_{\min})$, where E is the energy state for a particular strike, and E_{\min} and E_{\max} are the respective minimum and maximum energy state for that segment. Background shows bedrock topography with calderas outlined (in black). Grey thick line shows the edge of the Vatnajökull ice cap.

the strike of a new dyke segment will influence the strain and gravitational potential energy change in a different way. The direction that minimizes the combined energy should be favoured (Methods and Extended Data Figs 7 and 8). For the Bárðarbunga 2014 rifting event, the actual propagation path closely follows that predicted by our model (Fig. 3b) and can in particular explain why the dyke propagation changed to a northerly direction after initially propagating to the southeast. The influence of topography is large during the first segments but decreases as the dyke propagates towards more level topography and the tectonic stress becomes dominant in determining the direction of the dyke propagation. In essence, the dyke is captured by the plate spreading field once it is sufficiently far from the Bárðarbunga central volcano, which is located to the west of the central axis of the plate spreading model invoked (Methods). We have assumed in our model that the dyke remains at a fixed depth with respect to sea level, as it propagates. If in fact the dyke maintains a level of neutral buoyancy, the influence of topography will be about one-third greater (Methods), changing the predicted path slightly.

Our results show the dyke is heterogeneous in terms of seismic moment release and vertically integrated magma volume, peaking on the segments where the dyke halted, at 20–28 km and 33–39 km along the

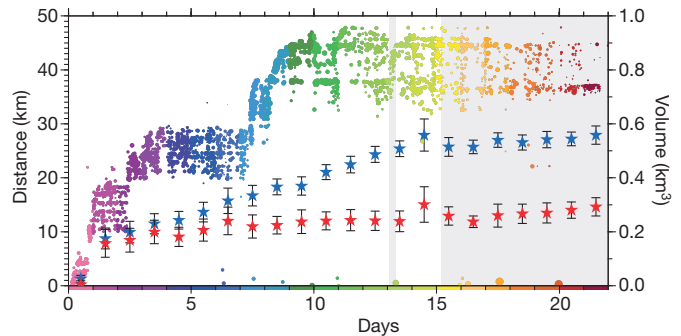


Figure 4 | Seismicity and volume change. Dots indicate earthquakes along the length of the dyke (left-hand y axis) as a function of time; stars, volume change (right-hand y axis) of the dyke (blue) and of the magma source (red). Earthquakes are colour coded by time (key along bottom), with dot radius proportional to earthquake magnitude. The volumes (error bars show 95% confidence intervals) are estimated from available geodetic data for each day using a model of a point pressure source and two dip-slip faults beneath the caldera. The magma source volumes are scaled by a factor of two, such that the value estimated for 5 September from GPS and InSAR data alone becomes equal to that estimated when the caldera subsidence is added to the inversion. Shading indicates the Holuhraun eruptions (see main text).

dyke (Fig. 2a). These are also locations where magma possibly reached the bedrock surface, as revealed by the ice cauldrons formed (Fig. 2b). The longest halt in the dyke propagation on 19–23 August correlates with increased lithostatic pressure, for any given depth, in the direction of propagation (Extended Data Fig. 8). Lateral dyke propagation is facilitated if a dyke advances into an area with falling lithostatic pressure, as the level of neutral buoyancy drops¹⁸. Such a process can be driven by gravity alone, but farther propagation when the lithostatic pressure increases requires the dyke to propagate upwards. Several days of magma flow to the Bárðarbunga dyke tip were required to increase the internal pressure sufficiently and drive propagation past the largest barrier along its path. Our seismic and geodetic observations provide details of a lateral dyke advance in segments, which can be related to the effects of the plate boundary stress field and topography on dyke steering and segmentation, with flow influenced by along-dyke variation in the lithostatic pressure profile. Similar studies, which may in future be carried out in near real-time, can lead to improved understanding of the evolution and forecasting of the behaviour of lateral dykes in various tectonic settings^{19,20}.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 8 October; accepted 25 November 2014.

Published online 15 December 2014.

- Wright, T. J. *et al.* Geophysical constraints on the dynamics of spreading centres from rifting episodes on land. *Nature Geosci.* **5**, 242–250 (2012).
- Einarsson, P. & Brandsdóttir, B. Seismological evidence for lateral magma intrusion during the July 1978 deflation of the Krafla volcano in NE-Iceland. *J. Geophys.* **47**, 160–165 (1980).
- Buck, W. R., Einarsson, P. & Brandsdóttir, B. Tectonic stress and magma chamber size as controls on dike propagation: constraints from the 1975–1984 Krafla rifting episode. *J. Geophys. Res.* **111**, B12404, <http://dx.doi.org/10.1029/2005JB003879> (2006).
- Belachew, M. *et al.* Comparison of dike intrusions in an incipient seafloor-spreading segment in Afar, Ethiopia: seismicity perspectives. *J. Geophys. Res.* **116** (B6), B06405, <http://dx.doi.org/10.1029/2010JB007908> (2011).
- Gudmundsson, A. Infrastructure and mechanics of volcanic systems in Iceland. *J. Volcanol. Geotherm. Res.* **64**, 1–22 (1995).
- Hartley, M. E. & Thordarson, T. The 1874–1876 volcano-tectonic episode at Askja, North Iceland: lateral flow revisited. *Geochem. Geophys. Geosyst.* **14**, 2286–2309 (2013).
- Segall, P. *Earthquake and Volcano Deformation* (Princeton Univ. Press, 2009).
- Segall, P., Llenos, A. L., Yun, S.-H., Bradley, A. M. & Syracuse, E. M. Time-dependent dike propagation from joint inversion of seismicity and deformation data. *J. Geophys. Res. Solid Earth* **118**, 5785–5804 (2013).

9. Björnsson, H. *Hydrology of Ice Caps in Volcanic Regions* Vol. 45 (Societas Scientiarum Islandica, 1988).
10. Gudmundsson, M. T. & Högnadóttir, T. Volcanic systems and calderas in the Vatnajökull region, central Iceland, constraints on crustal structure from gravity data. *J. Geodyn.* **43**, 153–169 (2007).
11. Björnsson, H. & Einarsson, P. Volcanoes beneath Vatnajökull, Iceland: evidence from radio-echo sounding, earthquakes and jökulhlaups. *Jökull* **40**, 147–168 (1990).
12. Thordarson, T. & Larsen, G. Volcanism in Iceland in historical time: volcano types, eruption styles and eruptive history. *J. Geodyn.* **43**, 118–152 (2007).
13. Einarsson, P. *et al.* Center of the Iceland hotspot experiences volcanic unrest. *Eos* **78**, 369–375 (1997).
14. Pagli, C. *et al.* Crustal deformation associated with the 1996 Gjalp subglacial eruption, Iceland: InSAR studies in affected areas adjacent to the Vatnajökull ice cap. *Earth Planet. Sci. Lett.* **259**, 24–33 (2007).
15. Reddy, J. N. *An Introduction to Continuum Mechanics* (Cambridge Univ. Press, 2013).
16. Dahm, T. Numerical simulations of the propagation path and arrest of fluid-filled fractures in the Earth. *Geophys. J. Int.* **141**, 623–638 (2000).
17. Maccaferri, F., Bonafede, M. & Rivalta, E. A quantitative study of the mechanisms governing dike propagation, dike arrest and sill formation. *J. Volcanol. Geotherm. Res.* **208**, 39–50 (2011).
18. Fialko, Y. A. & Rubin, A. What controls the along-strike slopes of volcanic rift zones? *J. Geophys. Res.* **104**, 20007–20020 (1999).
19. Toda, S., Stein, R. S. & Sagiya, T. Evidence from the AD 2000 Izu islands earthquake swarm that stressing rate governs seismicity. *Nature* **419**, 58–61 (2002).
20. Staudacher, T. *et al.* The April 2007 eruption and the Dolomieu crater collapse, two major events at Piton de la Fournaise (La Réunion Island, Indian Ocean). *J. Volcanol. Geotherm. Res.* **184**, 126–137 (2009).

Supplementary Information is available in the online version of the paper.

Acknowledgements Support for this work was received from the European Community's Seventh Framework Programme Grant No. 308377 (Project FUTUREVOLC), the Icelandic Research Fund (Project Volcano Anatomy), the Research Fund at the University of Iceland, NERC, the Geological Survey of Ireland and the US National Science Foundation (NSF). COSMO-SkyMed data were provided by the Italian Space Agency (ASI) and TerraSAR-X data by the German Space Agency (DLR) through the Icelandic Volcanoes Supersite project supported by the Committee on Earth Observing Satellites (CEOS). RADARSAT-2 data were provided by the Canadian Space Agency and MDA Corporation. Natural Resources Canada Earth Sciences Sector

contribution number 20140314. An intermediate TanDEM-X digital elevation model was provided by DLR under project IDEM_GEOL0123. We thank the following key persons for help with instrumentation and data: B. H. Bergsson, Þ. Jónsson, V. H. Kjartansson, S. Steinórsson, P. Erlendsson, H. Ólafsson, J. Söring and D. Craig. We also acknowledge the many others who have contributed to GPS, seismic and other field work in the study area. For GPS equipment and support, we acknowledge services provided by the UNAVCO Facility with support from the US NSF and National Aeronautics and Space Administration (NASA) under NSF Cooperative Agreements Nos EAR-0735156 and EAR-0711446. S. Jónsson (KAUST, Saudi Arabia) and T. Villemin (EDYTEM, Université de Savoie, France) also provided support to GPS. Seismic equipment: The British Geological Survey donated several of the broadband seismic sensors around Vatnajökull. We thank the SEISUK facility for loans to R.S.W. of seismometers under loan 980. Landsvirkjun contributed GPS instruments and seismic sensors north of Vatnajökull. The surveying aeroplane of Isavia (Icelandic Aviation Operation Services) mapped the subsidence of Bárðarbunga. The Icelandic Coast Guard provided aeroplane and helicopter support for field studies.

Author Contributions The writing of the paper, and the research it is based on, was coordinated by F.S., A.H., S.H., K.S.V., B.G.Ö. and other members of Icelandic Meteorological Office (IMO) seismic monitoring team. All authors contributed ideas and input to the research and writing of the paper. Modelling of geodetic data was done by A.H., E.R.H. and Th.A. Analysis and operation of continuous GPS sites were carried out by S.H., B.G.Ö., H.M.F., R.A.B., V.D., H.G. and P.C.L. Relative earthquake locations were calculated by K.S.V., seismic data presentation and relative locations were by G.B.G., focal mechanisms were determined by M.H. and single earthquake location determination was led by K.J. and SIL seismic monitoring group. Interferometric analysis was carried out by S.D., K.S., M.P., V.D. and S.S., with a composite digital elevation model prepared by E.M. The combined strain and gravity potential in relation to the dyke propagation was modelled by E.R.H. in consultation with A.H. and F.S. Campaign GPS measurements were carried out by S.D., V.D., M.P., Á.R.H., E.S., F.S. and others. M.T.G. and Th.H. mapped the collapse of the Bárðarbunga caldera and the formation of ice cauldrons over the path of the dyke with aircraft radar profiling. Á.R.H., E.M. and P.E. led mapping of graben formed and of the eruptive fissure as shown here. P.E., B.B. and R.P. contributed to the interpretation of events, and H.B. and F.P. provided bedrock topography and ice thickness for Vatnajökull ice cap. R.S.W., Th.Á., T.G., R.G.G., C.J.B., M.M., A.K.B. and E.P.S.E. contributed and analysed seismic data.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to F.S. (fs@hi.is).

METHODS

Seismic analysis. Seismicity was recorded by the Icelandic national seismic network (SIL) complemented with seismometer installations from the University of Cambridge and University College Dublin. Events attributed to the laterally growing dyke are volcano tectonic events. Initial single earthquake locations are performed by minimizing the square sum of both P- and S-wave arrival time residuals in the SIL analysis software²¹. Relative relocations are obtained by iterative inversion of the weighted square sums of: absolute P- and S-arrival time differences, as well as the double differences of (1) absolute arrival times of P and S waves, (2) relative arrival times of P and S waves and (3) relative S–P arrival times²². Each event is inverted in a group with over 40 of its nearest neighbours. Overlap of groups is enough to ensure that most events are located in at least five groups. The solutions shown are obtained using the SIL velocity model, which is the standard one-dimensional (1D) reference velocity model of the SIL analysis system²³. In the relative earthquake locations, the different elevations of seismic stations are not taken into consideration, except through the relative importance of the stations in the inversion for best locations (that is, the number of phases used). The average elevation of the dominant stations (0.9 km) is therefore taken as the initial reference elevation of the relative location results. To reference the location results to sea level, the depths were therefore shifted upwards by 0.9 km. To estimate the dependence of the location results on velocity model, the relative locations were also calculated in a second velocity model (IMO-vj), which is a rough 1D approximation to the velocity on the ICEMELT refraction profile at the northern margin of Bárðarbunga²⁴. This model gives source depths which are within 100 m in lateral distance but mostly around 2.5 km deeper than in the SIL model. This is probably caused by the lower velocities in the IMO-vj model below 6 km depth (Supplementary Fig. 7). Even though relative earthquake location errors can be quite small, there is always ambiguity about absolute location accuracies. The location of the ice depressions above the dyke segments where the dyke propagation temporarily stalled, and the location of the graben subsidence directly above the seismicity, confirm the quality of the absolute lateral locations. To further test the absolute depth accuracies, 100 events along the whole dyke were selected and located with NonLinLoc²⁵ in another approximation to the ICEMELT profile at Vatnajökull (CAM-vatnaj). The results, referenced to sea level (Supplementary Figs 7 and 8), show a very similar depth range to the relative locations and further support the absolute vertical location quality of the earthquakes. The two models, SIL and CAM-vatnaj, have a very different shallow structure, but below 6.5 km, where most of the seismicity is concentrated, they are very similar.

Focal mechanisms. These (Fig. 2) are best fitting solutions using a lower hemisphere projection based on grid search over all strike, dip and rake combinations matching observed P-wave polarities and within allowed limits from observed spectral amplitudes of P and S waves²⁶. Exemplary focal mechanisms of earthquakes $M > 2$ with at least six fitting P-wave polarities have been selected for each subcluster of the dyke intrusion. Mechanisms shown in Fig. 2 are rotated according to the directions of the rotated areas along the dyke. Focal mechanisms have tensional axes consistently orientated near perpendicular to the dyke as expected near dykes¹⁹, while pressure axes are variable depending on the location of the event with respect to the dyke (that is, above or in front of it).

GPS analysis. Significant deformation was observed at 16 pre-existing continuous GPS stations in relation to the Bárðarbunga events in 2014. Five additional sites were set up during the unrest leading up to the Holuhraun fissure eruption, all installed next to or on existing monuments (Supplementary Table 1). An additional 16 sites were measured regularly during the unrest (Supplementary Table 2). Multiple measurements were made at all these sites before the Bárðarbunga unrest, with the exception of site GSIG, which was installed and first measured in June 2014. GSIG is located about 700 m from an existing benchmark. The last pre-unrest GPS campaign in the region was conducted from 28 July to 9 August 2014.

The GPS data were analysed using the GAMIT/GLOBK software, version 10.4 (ref. 27), using over 100 global reference stations. Average site positions were evaluated in the ITRF08 reference frame every 24 h UTC day. The continuous GPS data were furthermore divided into three eight-hour sessions with a running 24 h window of reference station and orbit data, to provide higher temporal resolution (Supplementary Fig. 1a–f). In addition to station coordinates, the processing solved for satellite orbit and Earth rotation parameters, atmospheric zenith delay every two hours, and three atmospheric gradients per day. Ocean loading was corrected for using the FES2004 model. The IGS08 azimuth and elevation dependent absolute phase centre model was applied for all antennas. Pre-rifting site velocities were estimated based on all existing data and removed from the data. The last three to six days of measurements at each site before 16 August were then used to estimate a reference epoch. Data affected by snow and ice were removed during the analysis. **Interferometric analysis.** This analysis of X-band satellite data (wavelength ~ 3.1 cm) from the COSMO-SkyMed and TerraSAR-X satellites was undertaken using the Repeat Orbit Interferometry Package (ROI_PAC)²⁸ and DORIS software²⁹.

Analysis of C-band RADARSAT-2 data (wavelength ~ 5.56 cm) was computed using the GAMMA software³⁰. Topographic signal in the interferograms was estimated using a LiDAR DEM³¹ on the glacier and for an area extending 2–3 km from the glacier margin. Further from the glacier, an intermediate DEM from the TanDEM-X mission was used with a DEM from the ASTER satellite mission and the EMISAR DEM³² to fill in observed gaps. The DEM mosaic used for the topographic correction has a pixel size of about $13 \text{ m} \times 30 \text{ m}$ (the pixel size of the ASTER DEM), at the latitude of Iceland. Interferograms were filtered using a power spectrum filter³³ and unwrapped using the branch cut algorithm³⁴ and the SNAPHU minimum-cost-flow method³⁵. We downsampled all interferograms using an adaptive quad-tree approach³⁶, with a cut-off variance of 10^{-3} m^2 . Interferograms are shown in Supplementary Fig. 2 and a list of all interferograms used is in Supplementary Table 3.

The subsiding graben. This was mapped using high resolution radar images from the airborne radar system on board the Icelandic Coast Guard aircraft TF-SIF and photographs taken on board the same aeroplane. The photographs were also used to obtain coordinates for the eruptive fissures. The photographs were geo-referenced by comparison with older geo-referenced aerial photographs from Loftmyndir Corporation, using ArcGIS software. The radar images were geo-referenced with the LiDAR DEM³¹ using MATLAB R2013a and Surfer 12 (Golden Software, Inc.).

Deformation modelling. Measurement errors were assumed to be drawn from a zero-mean Gaussian distribution and the errors in the physical model were assumed to scale up the effective measurement error. Application of Bayes' theorem gives the a posteriori probability distribution for the model parameters as

$$p(\mathbf{m}, \sigma, \mathbf{d}) = K(\sigma^2)^{-N/2} \exp \left[-\frac{1}{2\sigma^2} (\mathbf{d} - \mathbf{Gm})^T \Sigma_d^{-1} (\mathbf{d} - \mathbf{Gm}) \right] p(\mathbf{m})$$

where \mathbf{m} is the vector of model parameters, \mathbf{d} is the vector of measurements, \mathbf{G} is a matrix of Green's functions mapping slip to displacements, Σ_d is the variance-covariance matrix for the measurements, σ^2 is the scaling factor due to model error, N is the number of measurements, K is a normalizing constant, and $p(\mathbf{m})$ is the a priori probability of the model parameters. The covariance of the error for each pair of InSAR measurements is calculated assuming a 1D exponential covariance function: $\text{Cov} = 0.0016 \exp(-h/5) \text{ m}^2$, where h is the distance between the measurement points in km. The model parameters are opening and strike-slip for the dyke patches³⁷, position and pressure decrease of a penny-shaped crack³⁸ or point pressure source³⁹, a bilinear orbital error ramp for each interferogram, and the hyperparameter σ^2 . We allow for slip as well as opening, as dykes that are not perpendicular to the minimum compressive stress direction will be subject to shearing across the dyke walls⁴⁰. We set the a priori probability to allow only positive opening and slip in the direction consistent with the regional stress field from relative plate motions. During the geodetic modelling the different elevation of geodetic stations was not taken into consideration. The initially inferred depths were therefore shifted by the average elevation of the GPS stations (1.0 km), resulting in geodetic model depths relative to sea level (shown here).

The a posteriori distribution is sampled using a Markov chain Monte Carlo algorithm, incorporating the Metropolis algorithm⁴¹. This involves selecting an initial value for each of the model parameters from $p(\mathbf{m})$ and calculating the likelihood function, which is the right-hand side of the equation above excluding $p(\mathbf{m})$. A trial random step is then taken within $p(\mathbf{m})$, and the new likelihood value is calculated. If the new likelihood value is greater, the step is taken and the trial model values are retained. If less, there is a chance that the step is taken, which is calculated as the ratio of the new likelihood over the old likelihood. Otherwise the old model values are retained. A new trial random step is taken, and the process is repeated until a representative sampling of the whole a posteriori distribution is built. The efficiency of this algorithm in reaching this goal depends on the maximum size of the random step that may be taken within $p(\mathbf{m})$. In order to ensure fast convergence, we perform a sensitivity test for each model parameter after every 1,000 iterations, and adjust the maximum step size such that all parameters contribute approximately equally to the change in likelihood and, as a whole, the mean chance of acceptance is approximately 50% (ref. 42).

The models of the deflation at Bárðarbunga are more uncertain than the dyke; however, whichever model we choose for the deflation, the modelled values of dyke opening do not change significantly.

Strain potential energy change. To calculate this change, associated with dyke formation, we require an estimate of tectonic stress (deviatoric stress induced by plate movements). To estimate the strain potential we assumed that the tectonic stress due to plate spreading could be estimated by an infinitely long and wide tensile dislocation below 10 km depth in an elastic half-space. Such a kinematic model has been used successfully to fit GPS observations across the plate boundary in Iceland⁴³. This tensile dislocation was opened 4 m, which would correspond to stress built up by plate spreading for more than 200 years. It was located so that it would be under the Askja central volcano as geodetic measurements have indicated that the central

axis of plate spreading passes through there⁴⁴. The strike of this dislocation was set to N12°E, to be about perpendicular to direction of plate movements predicted by global plate motion models. We assume the tectonic stress throughout the depth interval of the crust considered does not vary with depth, similarly to the approach of Buck *et al.*³. The value of stress we use is that calculated at 10 m depth in the dislocation model. We calculate the stress and strain due to a dyke segment opening in a similar manner and superpose them on the estimated tectonic contributions. Assuming a linear relationship between stress and strain, we then calculate the strain energy potential using equation (2).

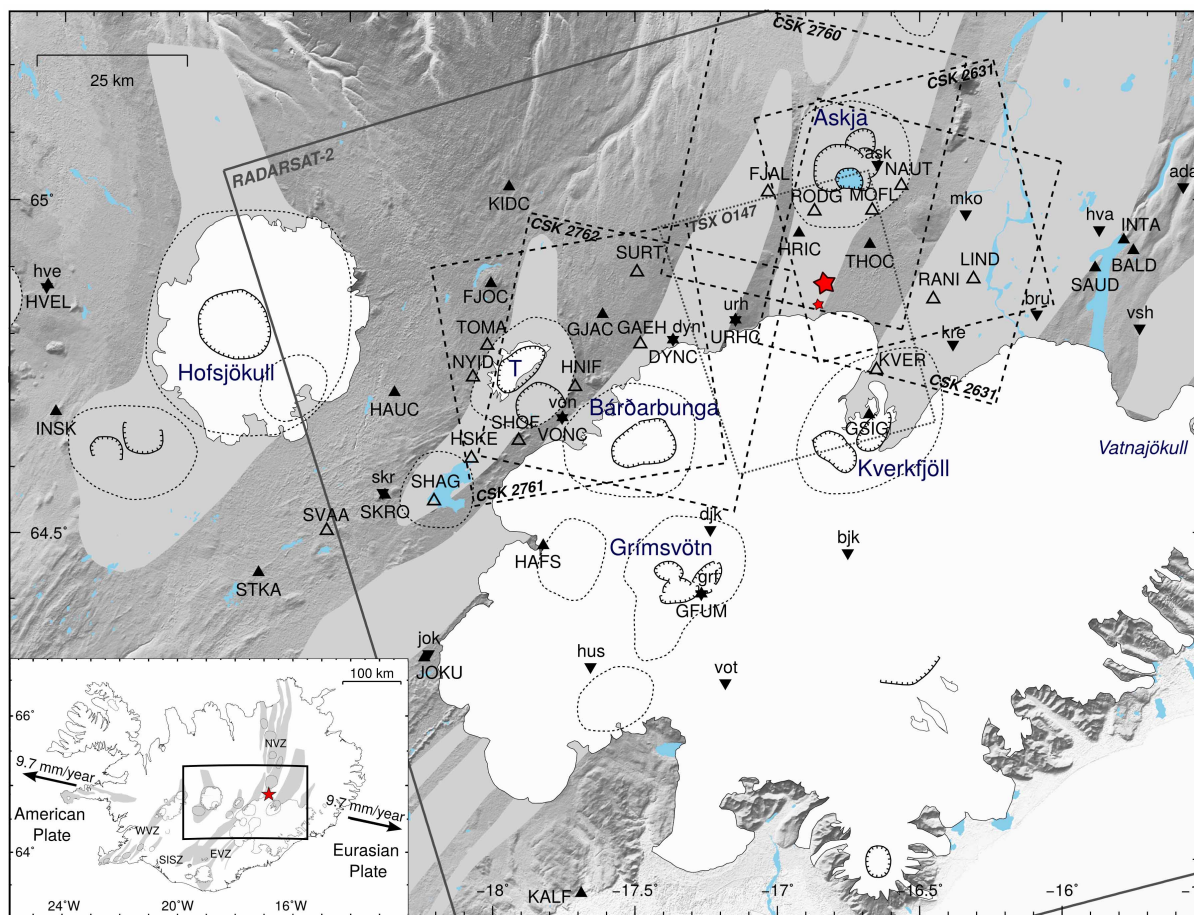
Gravitational potential energy change. This quantity is here calculated, for each dyke segment, as described in the text by integrating the predicted vertical displacements associated with the dyking, multiplied by the local topographic load density (ice and crust) above an arbitrary reference surface (taken here as sea level) and the acceleration of gravity. Two digital elevation models are used, one of which covers the surface of the Vatnajökull ice cap and extends beyond the limits of the ice cap, and the other which represents the ice thickness. The map of the sub-ice topography was compiled from continuous ice thickness profiling by radio echo-sounding along a series of traverses over the ice cap^{9,11}. Along the complete length of the dyke the change in lithostatic pressure corresponds to an effective crustal load change of about 900 m (Extended Data Fig. 8). However, variations in the effective load in areas adjacent to an individual dyke segment influenced by vertical displacements are much smaller, typically on the order of several hundred metres or less. This is an order of magnitude less than the 2 km depth to the top of a 'test dyke segment' used for calculation of the preferred path of dyking (see below). Thus, we can consider small perturbations to the vertical deformation field introduced by the real topography to be second order.

When inferring the path of preferred dyke propagation, we assume also that the dyke depth, with respect to sea level, is the same for all strikes tested. In reality the dyke may track the level of neutral buoyancy, resulting in the preferred depth of dyking varying with strike. In our approach, the dyke moving down by one metre (with respect to sea level) can be considered equivalent to increasing the load on the reference surface by one metre of crust. The associated increase in potential energy change (compared to that when the dyke stays at the same depth) will be equal to the integrated vertical displacement of the reference surface multiplied by the density of the crust and gravitational acceleration. On the other hand, the reduction in potential energy from lowering the magma will be equal to the volume of the dyke multiplied by the density of the magma and gravitational acceleration. For a Poisson's ratio of 0.25, the integrated surface uplift is 75% of the dyke volume⁴⁵. Therefore ~25% less energy is needed to lift the extra crust than is released by lowering the dyke. That is, the energy released in lowering the dyke is ~33% more than needed to lift the extra crust. This means that if the dyke propagates at a level of neutral buoyancy, rather than remaining at a fixed depth (with respect to sea level), the differences in gravitational potential energy change with strike will be ~33% larger than we calculate, thus increasing the influence of topography still further.

Calculation of the preferred path of dyking. For combined potential energy change during dyking, we here estimate all parameters based on seismic and geodetic data except the strike of a dyke segment. Each segment, whose location and length are determined from relative earthquake locations, is assumed to be a rectangular tensile dislocation³⁶. The depth to the top of each dislocation is fixed to 2 km for all segments, the width (height) is fixed to 4 km and opening is fixed to 3 m. The starting point of each segment is fixed adjacent to the previous segment (black dots on Fig. 3b). This ensures that only energy states which assume continuation of the magma flow are considered. Then the strike of the segment is varied so that it is rotated around its starting point. The strike is varied well over 180° in search of the minimum energy for emplacement of the new segment. To implement the approach we performed two integrations, one in three dimensions for the strain potential energy and one in two dimensions for the gravitational potential energy. A Monte Carlo numerical integration in MATLAB was used, where a mean value was estimated and multiplied by the volume for strain energy, or area for gravitational energy. The rectangular dislocation formula does not take into account the strength of the material and in its vicinity the strain energy density is non-realistic and close to singular values. To avoid these values we assigned zero to energy density values over three orders of magnitude larger than the estimated average value. Therefore, we did not evaluate the strain energy densities in the immediate vicinity of a dyke intrusion, but rather evaluated how tectonically stressed crust will respond to dyke opening and if that opening will increase or decrease the total potential energy of the crust. The area of integration included a radius greater than 50 km

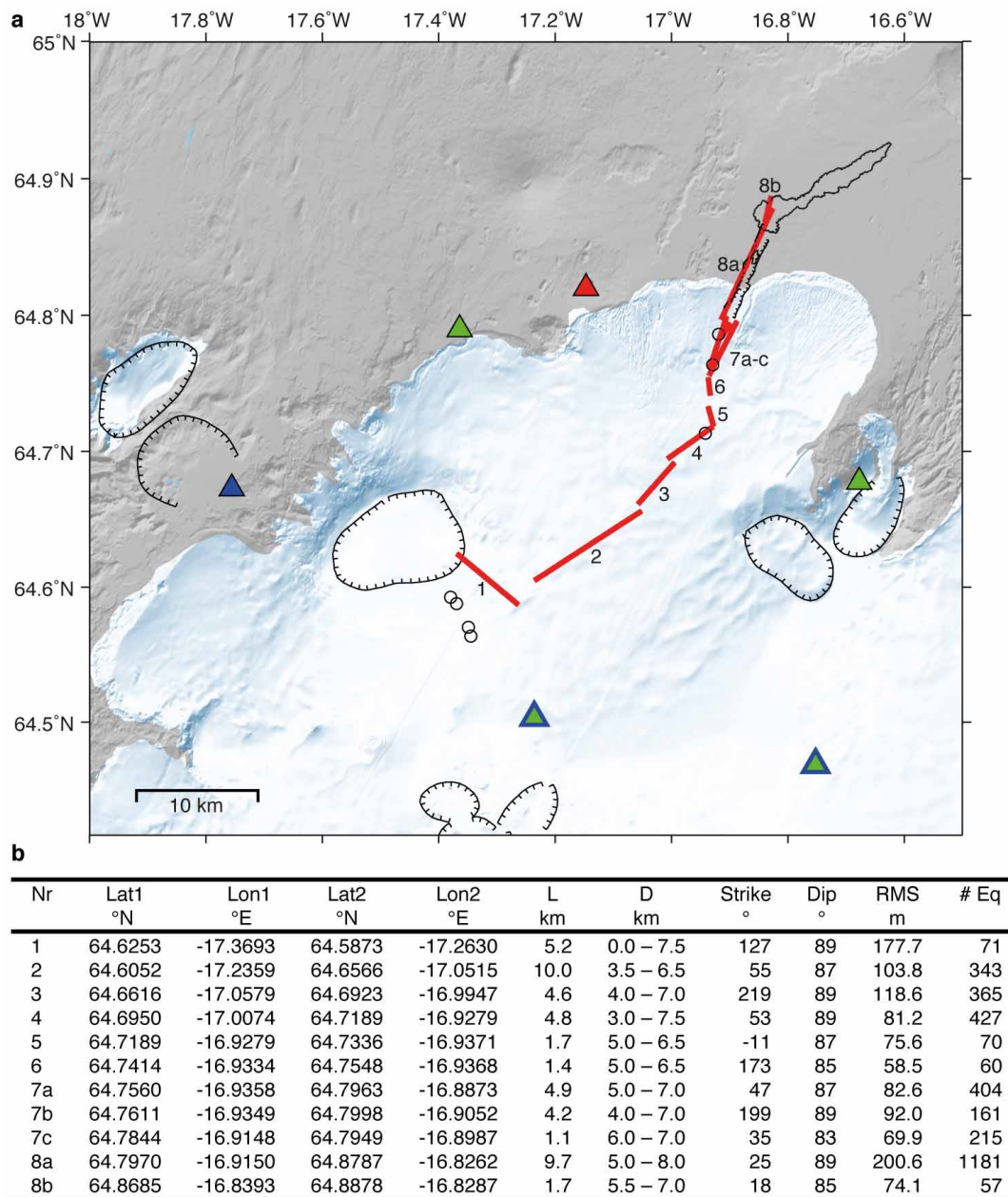
from each dyke segment. The strain energy density was integrated down to a depth of 20 km. We found that this was sufficient so that the boundaries did not influence the estimated energy changes. To calculate the stress and strain tensors as well as the vertical surface displacements we used *disloc3d*, software developed by the Crustal Deformation and Fault Mechanics research group at Stanford University.

21. Bödvarsson, R., Rögnvaldsson, S. T., Slunga, R. & Kjartansson, E. *The SIL Data Acquisition System — At Present and Beyond Year 2000* (Report VI-R98005-JA04, Icelandic Meteorological Office, 1998).
22. Slunga, R., Rögnvaldsson, S. T. & Bödvarsson, R. Absolute and relative locations of similar events with application to microearthquakes in southern Iceland. *Geophys. J. Int.* **123**, 409–419 (1995).
23. Stefánsson, R. *et al.* Earthquake prediction research in the South Iceland seismic zone and the SIL project. *Bull. Seismol. Soc. Am.* **83**, 696–716 (1993).
24. Darbyshire, F. A., Bjarnason, I. Th., White, R. S. & Flóvenz, Ó. G. Crustal structure above the Iceland mantle plume imaged by the ICEMELT refraction profile. *Geophys. J. Int.* **135**, 1131–1149 (1998).
25. Lomax, A., Virieux, J. J., Volant, P. & Berge, C. in *Advances in Seismic Event Location* (eds Thurber, C. H. & Rabinowitz, N.) 101–134 (Kluwer, Amsterdam, 2000).
26. Rögnvaldsson, S. T. & Slunga, R. R. Routine fault plane solutions for local networks: a test with synthetic data. *Bull. Seismol. Soc. Am.* **83**, 1232–1247 (1993).
27. Herring, T., King, R. W. & McCluskey, S. M. *Introduction to GAMIT/GLOBK Release 10.4* (Massachusetts Institute of Technology Technical Report, Cambridge, 2010).
28. Rosen, P. A., Henley, S., Peltzer, G. & Simons, M. Updated repeat orbit interferometry package released. *Eos Trans. AGU* **85**, 47 (2004).
29. Kampes, B. *Delft Object-oriented Radar Interferometric Software: User's Manual and Technical Documentation v4.02* (Delft Technical University, 1999).
30. Werner, C. *et al.* Gamma SAR and interferometric processing software. *Proc. ERS-ENVISAT Symp.* Vol. 461 (European Space Agency Special Publication, 2000).
31. Jóhannesson, T. *et al.* Ice-volume changes, bias-estimation of mass-balance measurements and changes in subglacial water bodies derived by LiDAR-mapping of the surface of Icelandic glaciers. *Ann. Glaciol.* **54**, 63–74 (2013).
32. Magnússon, E., Björnsson, H., Dall, J. & Pálsson, F. Volume changes of Vatnajökull ice cap, Iceland, due to surface mass balance, ice flow, and subglacial melting at geothermal areas. *Geophys. Res. Lett.* **32**, L05504, <http://dx.doi.org/10.1029/2004GL021615> (2005).
33. Goldstein, R. R. & Werner, C. Radar interferogram filtering for geophysical applications. *Geophys. Res. Lett.* **25**, 4035–4038 (1998).
34. Goldstein, R., Zebker, H. & Werner, C. Satellite radar interferometry-two-dimensional phase unwrapping. *Radio Sci.* **23**, 713–720 (1988).
35. Chen, C. W. & Zebker, H. A. Two-dimensional phase unwrapping with use of statistical models for cost functions in nonlinear optimization. *J. Opt. Soc. Am.* **18**, 338–351 (2001).
36. Decriem, J. *et al.* The 2008 May 29 earthquake doublet in SW Iceland. *Geophys. J. Int.* **181**, 1128–1146 (2010).
37. Okada, Y. Internal deformation due to shear and tensile faults in a half-space. *Bull. Seismol. Soc. Am.* **82**, 1018–1040 (1992).
38. Fialko, Y., Khazan, Y. & Simons, M. Deformation due to a pressurized horizontal circular crack in an elastic half-space, with applications to volcano geodesy. *Geophys. J. Int.* **146**, 181–190 (2001).
39. Mogi, K. Relations between the eruptions of various volcanoes and the deformations of the ground surfaces around them. *Bull. Earthq. Res. Inst. Univ. Tokyo* **36**, 99–134 (1958).
40. Hooper, A. *et al.* Increased capture of magma in the crust promoted by ice-cap retreat in Iceland. *Nature Geosci.* **4**, 783–786 (2011).
41. Mosegaard, K. & Tarantola, A. Monte Carlo sampling of solutions to inverse problems. *J. Geophys. Res.* **100**, 12431–12447 (1995).
42. Hooper, A. *et al.* Importance of horizontal seafloor motion on tsunami height for the 2011 $M_w=9.0$ Tohoku-Oki earthquake. *Earth Planet. Sci. Lett.* **361**, 469–479 (2013).
43. Árnadóttir, T., Jiang, W., Feigl, K. L., Geirsson, H. & Sturkell, E. Kinematic models of plate boundary deformation in southwest Iceland derived from GPS observations. *J. Geophys. Res.* **111**, B07402, <http://dx.doi.org/10.1029/2005JB003907> (2006).
44. Sturkell, E. & Sigmundsson, F. Continuous deflation of the Askja caldera Iceland, during the 1983–1998 non-eruptive period. *J. Geophys. Res.* **105**, 25671–25684 (2000).
45. Delaney, P. T. & McTigue, D. F. Volume of magma accumulation or withdrawal estimated from surface uplift or subsidence, with application to the 1960 collapse of Kilauea volcano. *Bull. Volcanol.* **56**, 417–424 (1994).
46. Jóhannesson, H. & Sæmundsson, K. *Geological Map of Iceland, Tectonics 1:600,000* (Icelandic Institute of Natural History, 2009).
47. Gudmundsson, M. T., Högnadóttir, Þ., Kristinsson, A. B. & Gudbjörnsson, S. Geothermal activity in the subglacial Katla caldera, Iceland, 1999–2005, studied with radar altimetry. *Ann. Glaciol.* **45**, 66–72 (2007).
48. Yun, S., Segall, P. & Zebker, H. Constraints on magma chamber geometry at Sierra Negra Volcano, Galápagos Islands, based on InSAR observations. *J. Volcanol. Geotherm. Res.* **150**, 232–243 (2006).



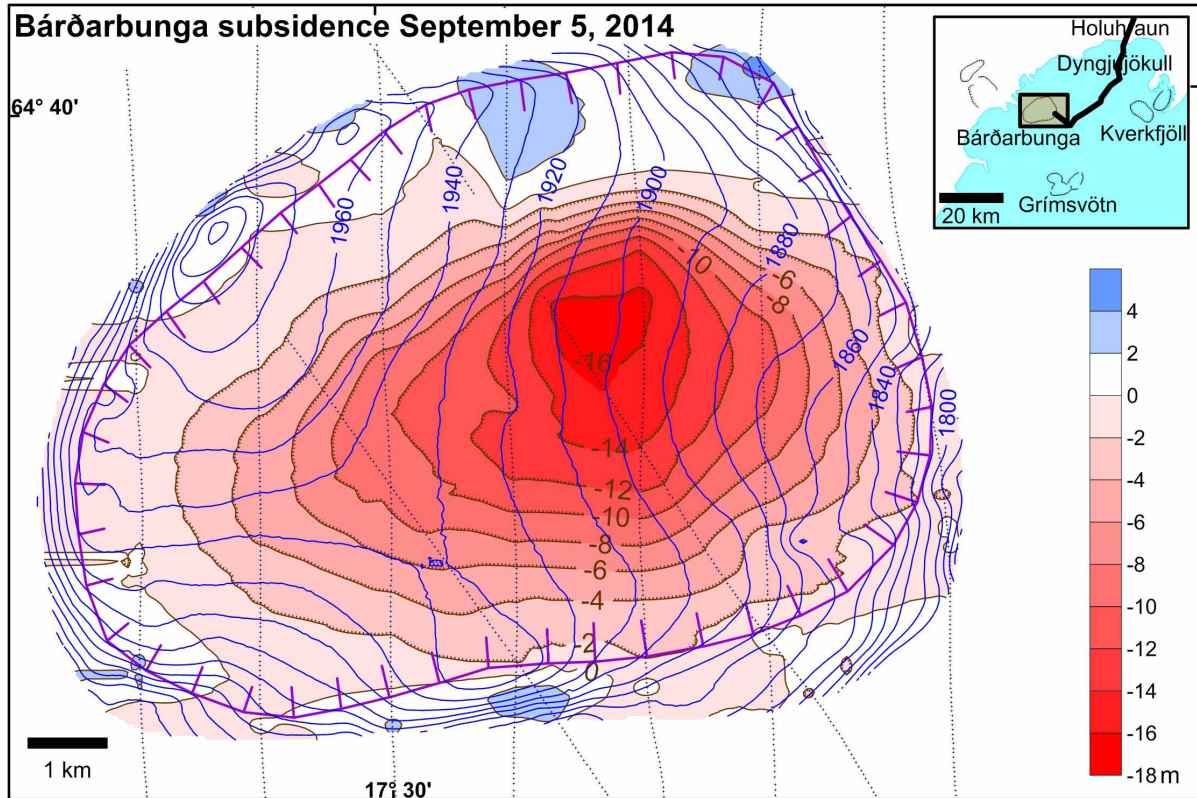
Extended Data Figure 1 | Tectonic map showing seismic and geodetic stations. Filled uptriangles correspond to continuous GPS stations, open triangles to campaign GPS sites, filled downtriangles correspond to seismic stations, and black 'stars' correspond to co-located GPS and seismic stations. Station names for GPS are indicated with four upper-case letters and for seismic stations with three lower-case letters. The different tracks for SAR satellite data are plotted with solid lines for RADARSAT-2, dashed lines for COSMO-SkyMed (CSK) and dotted lines for TerraSAR-X (TSX). Orbit numbers are

indicated for CSK and TSX. The red stars correspond to the 2014 eruptive fissures at Holuhraun. Background map shows ice caps (white), central volcanoes (dotted lines), calderas (hatched lines) and fissure swarms (grey shading)⁴⁶. Names of selected volcanoes are shown, and T indicates Tungnafellsjökull central volcano. Inset, locations of the Eastern Volcanic Zone (EVZ), the Western Volcanic Zone (WVZ), the Northern Volcanic Zone (NVZ) and the South Iceland Seismic Zone (SISZ), with their fissure swarms and central volcanoes; box shows area of the main figure.



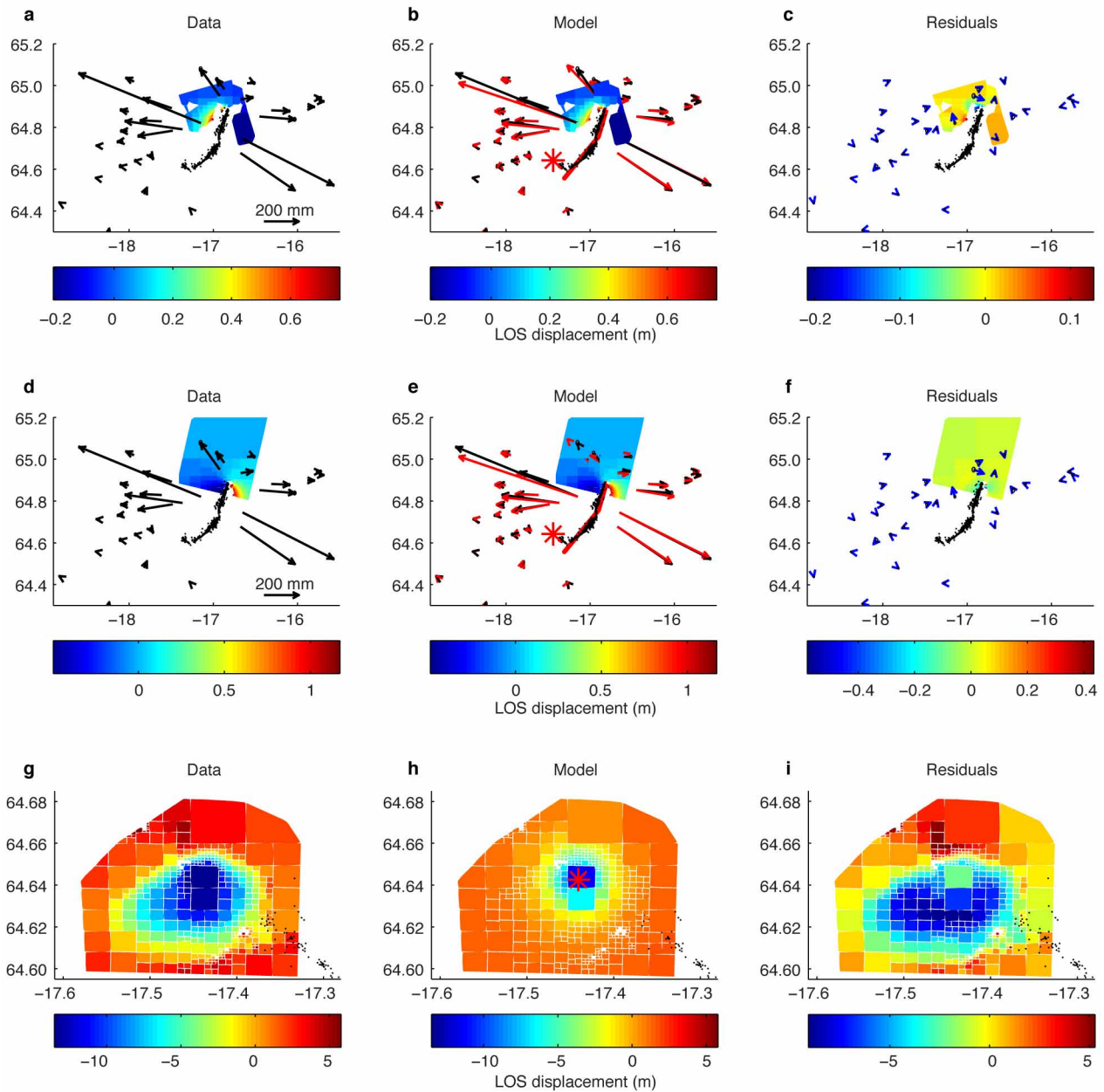
Extended Data Figure 2 | Map and table of dyke segments defined by seismicity. **a**, Location of dyke segments (red lines labelled with numbers) delineated by relatively relocated earthquakes. The triangles show locations of the nearest seismic stations used to locate the events. Green, stations operated by the Icelandic Meteorological Office (IMO); blue, station operated by the University of Cambridge; and red, station operated by University College Dublin. The two stations on the ice are joint IMO and University of Cambridge stations. All stations telemetered data to IMO. Also shown are ice cauldrons

formed (circles), outline of lava flow mapped from radar image on 6 September, and boundaries of graben activated in the dyke distal area (hatched lines). **b**, The dyke segments. Columns show segment number (Nr), latitude and longitude of beginning (Lat1, Lon1) and end points (Lat2, Lon2) of each segment, segment length (*L*), depth range (*D*), strike, dip, the RMS value of the deviation (in metres) of the earthquakes from the plane they define, and the number of earthquakes used to define each dyke segment plane (#Eq).



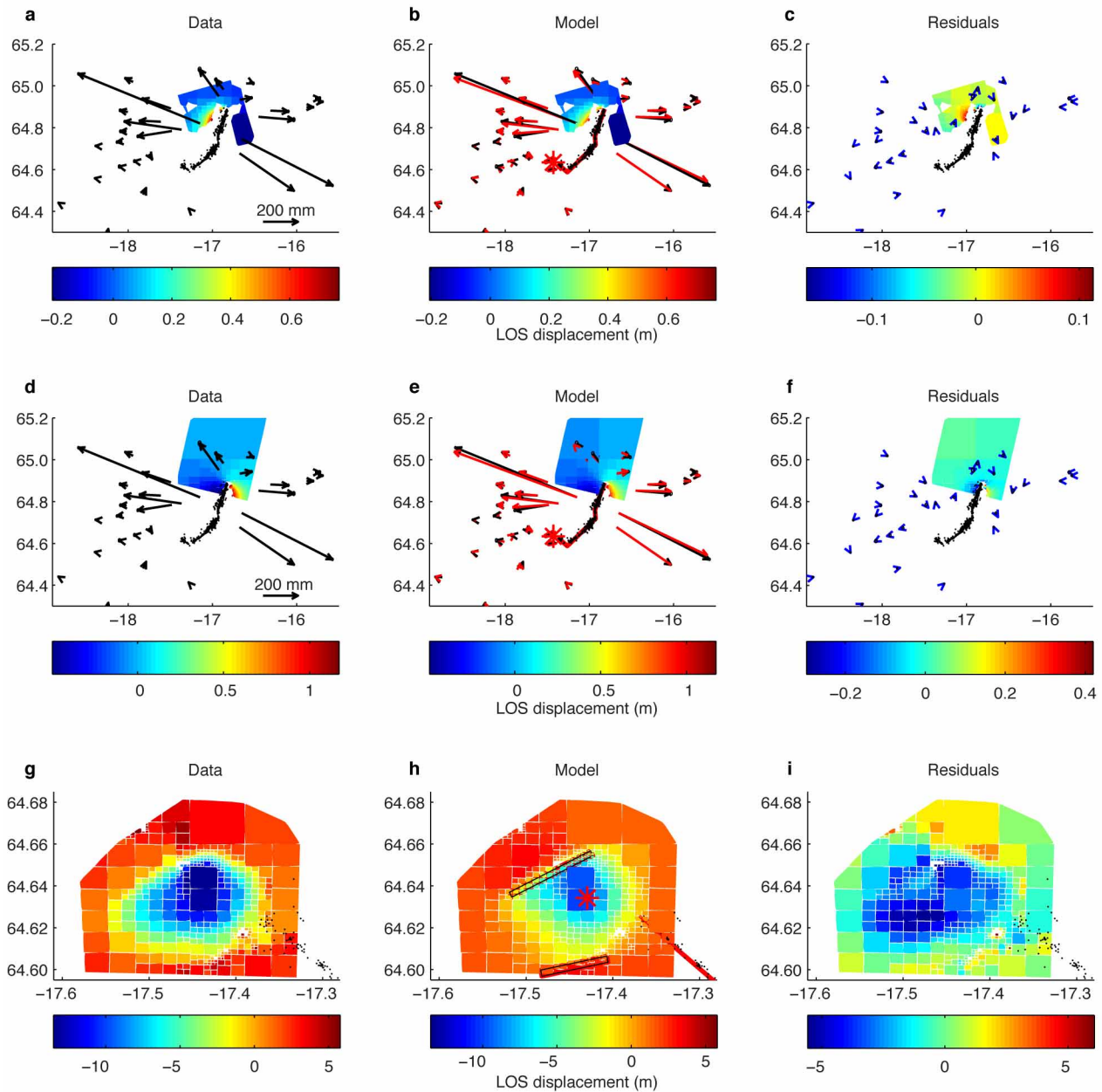
Extended Data Figure 3 | Subsidence at Bárðarbunga volcano revealed by aeroplane radar profiling. Blue contours show the elevation of the ice surface in the Bárðarbunga caldera on 5 September 2014, about three weeks after the onset of unrest. The data were obtained using aircraft flown 100–150 m above glacier surface, using radar altimetry and submetre differential GPS

Omnistar. The system provides 2 m absolute accuracy of surface elevation along the survey profiles⁴⁷, shown as black dotted lines. The subsidence relative to the pre-unrest ice surface is indicated with coloured shading (key at right). It is greatest in the central part of the caldera, where it had a maximum of 16 m.



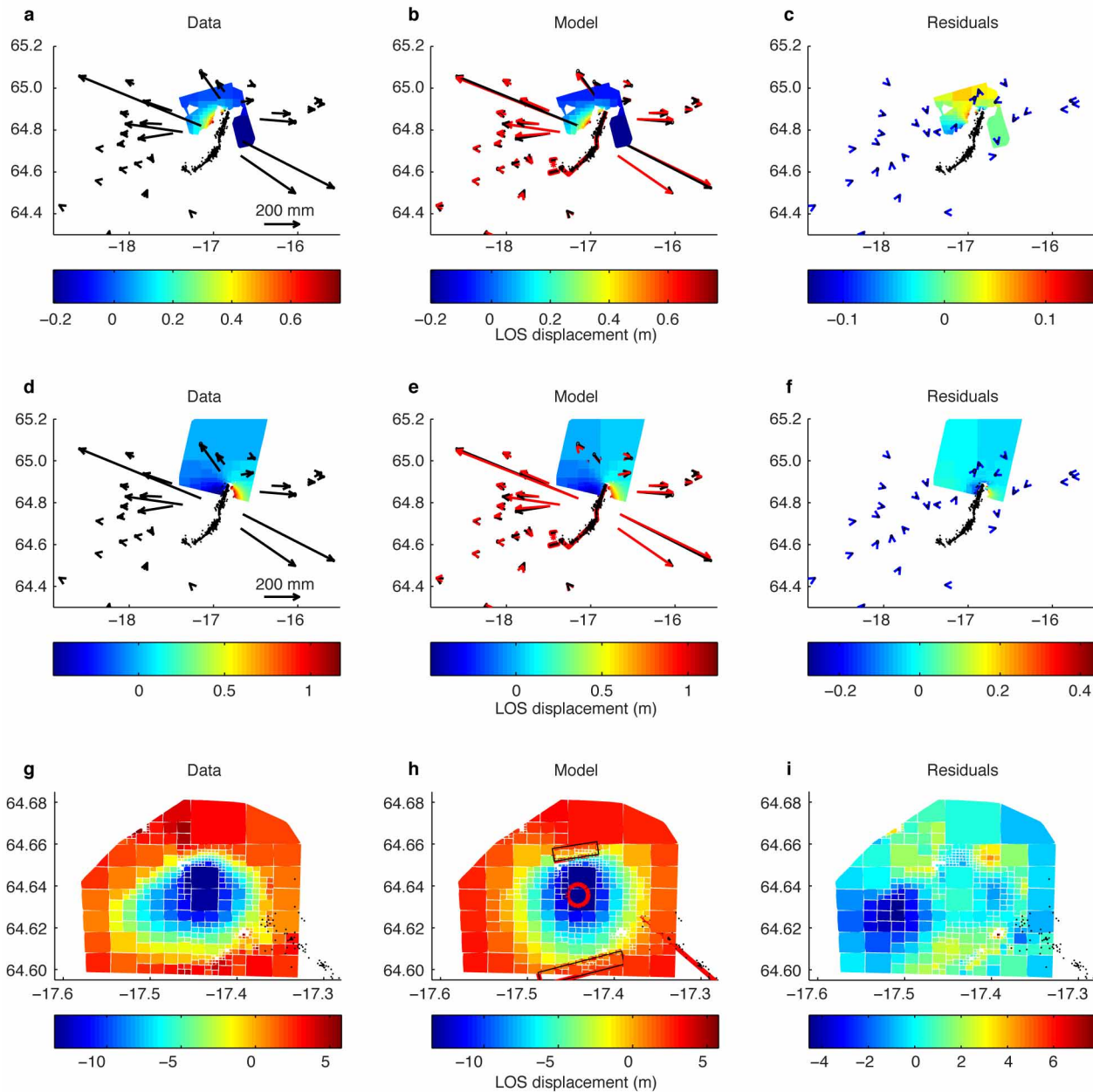
Extended Data Figure 4 | Geodetic model with a contracting point pressure source, a two-segment-dyke and no a priori constraints. The dyke segments are rectangular with uniform opening³⁷ and the model shown is that with maximum probability. Free parameters are: position and volume change of the point pressure source³⁹, and position, strike, dip and opening of the dyke. The panels show from left to right: data (a, d, g), model (b, e, h) and residuals

(c, f, i). GPS data in all panels span 15 August to 4 September 2014. The top row shows an interferogram spanning 6 July 2012 to 4 September 2014; the middle row shows an interferogram spanning 2 August to 3 September 2014; the bottom row shows the data from Extended Data Fig. 3 from aeroplane radar profiling, but these data were not used to constrain the model.



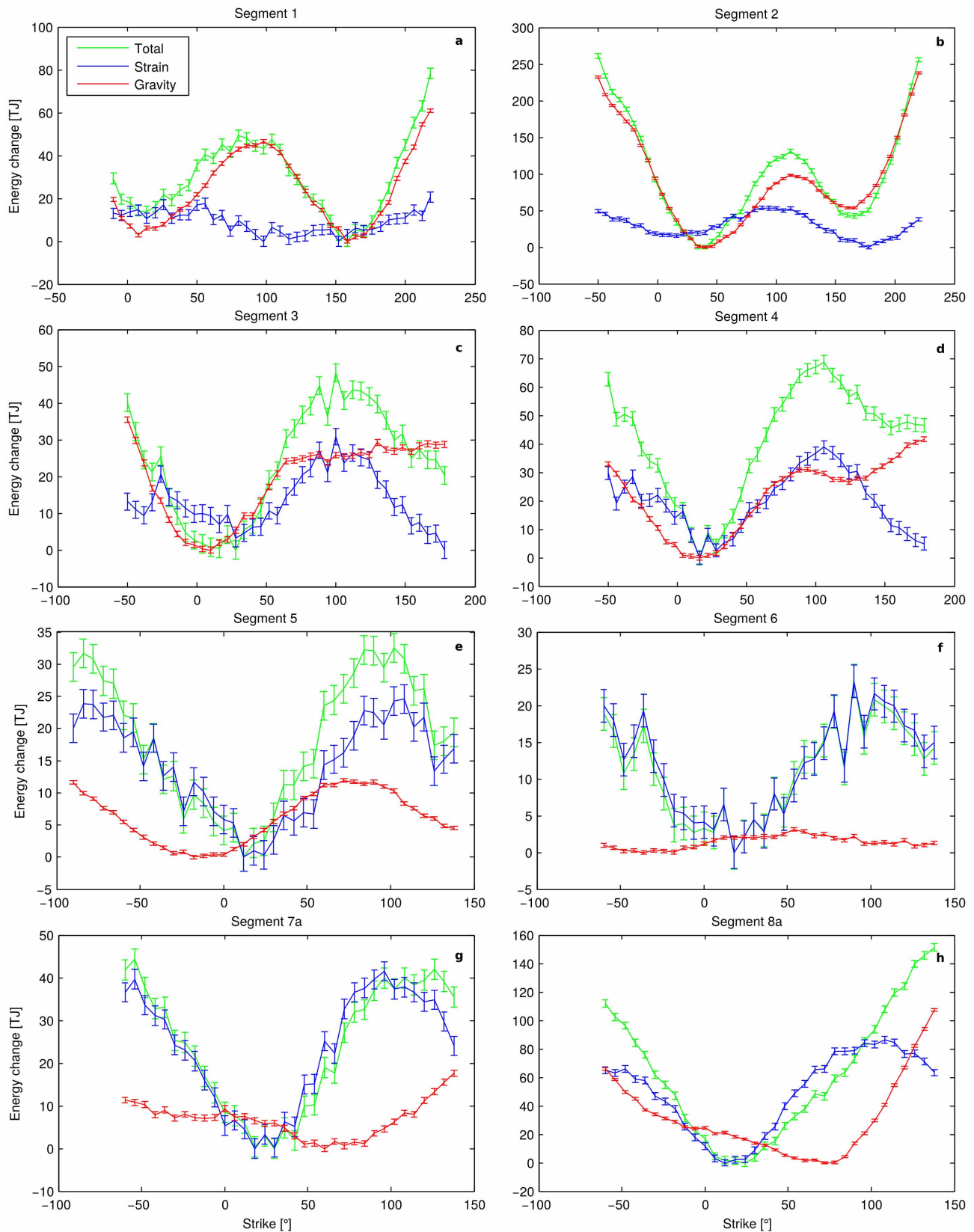
Extended Data Figure 5 | Geodetic model with a contracting point pressure source, two caldera faults, and a four-segment dyke. The dyke is divided into multiple rectangular patches³⁷ and the model shown is that with maximum probability. Lateral dyke position is fixed from relocated seismicity. Position of the point source and faults, volume change of the point source and opening

and strike-slip of the dyke are free parameters. Left column (a, d, g), data; middle column (b, e, h), model; right column (c, f, i), residuals. The data are detailed in Extended Data Fig. 4 and all of the data were used to constrain the model.



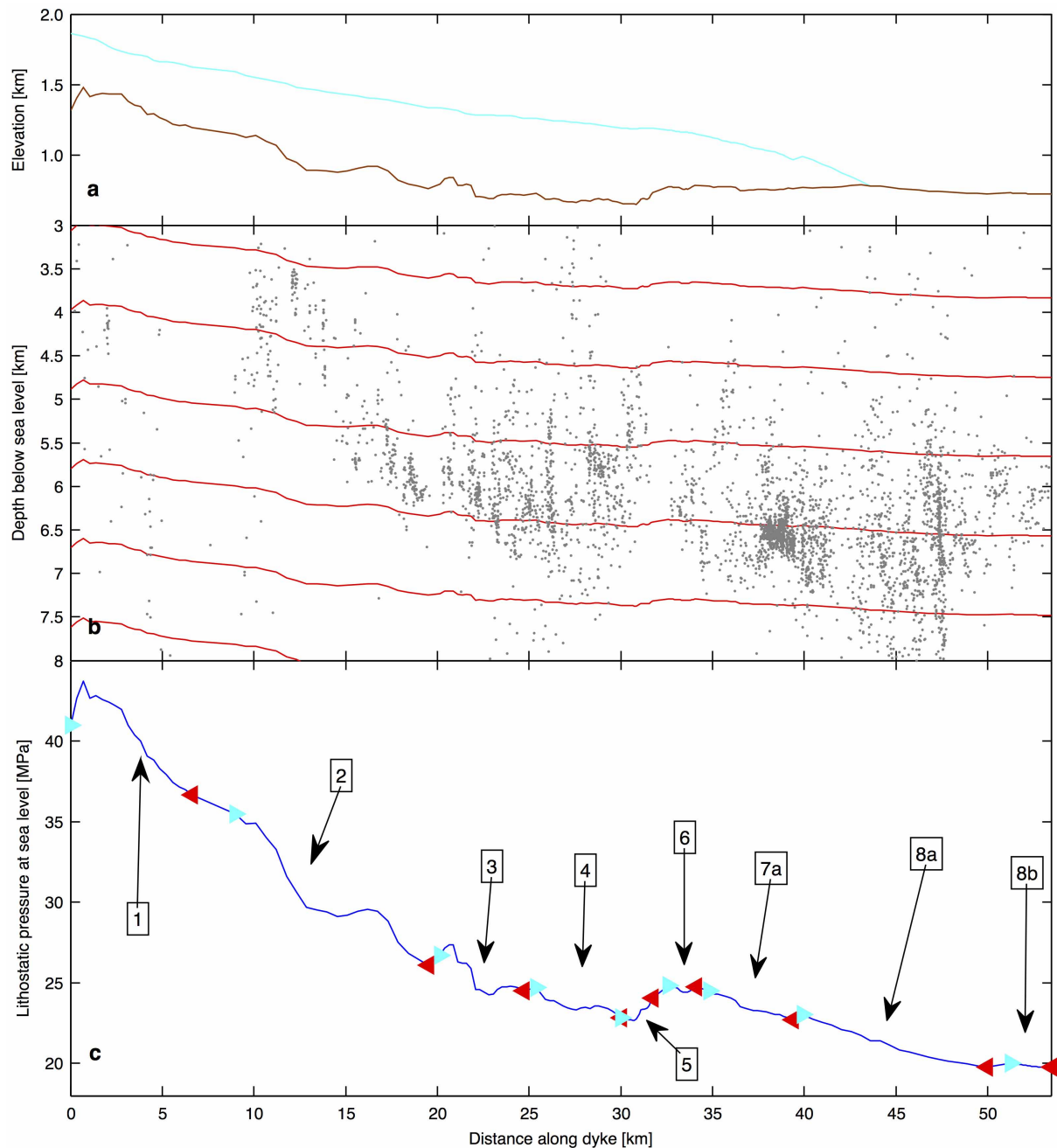
Extended Data Figure 6 | Geodetic model with a contracting flat-topped chamber, two caldera faults, and a four-segment dyke. A deflating penny-shaped crack³⁸ is used to represent the top of a flat-topped chamber⁴⁸, and the dyke is divided into multiple rectangular patches³⁷. The model shown is that with maximum probability. Lateral dyke position is fixed from relocated

seismicity. Position of the crack and faults, volume change and radius of the crack and opening and strike-slip of the dyke are free parameters. Left column (a, d, g), data; middle column (b, e, h), model; right column (c, f, i), residuals. The data are detailed in Extended Data Fig. 4 and all of the data were used to constrain the model.



Extended Data Figure 7 | Path of dyke propagation from energy considerations. a–h, Energy profiles for dyke segments 1, 2, 3, 4, 5, 6, 7a and 8a, as described in Fig. 3b. Blue lines indicate the strain energy potential change as a function of the strike, and the red lines the gravitational potential

change. Green is the total potential energy change. Energy is shown in terajoules (10^{12} J). The lowest point on each energy curve is defined as 0 TJ. Error bars, 1 s.d. of the error in the numerical integration.



Extended Data Figure 8 | Topography, earthquake depths, and lithostatic pressure along the dyke path. **a**, Bedrock (brown line) and ice topography (light blue line) along the dyke path. **b**, Depth of earthquake hypocentres (grey dots) below sea level projected on the dyke segments and lines (red) of constant lithostatic pressure, assuming constant crustal density of $2,800 \text{ kg m}^{-3}$ and ice density of 920 kg m^{-3} . Line spacing corresponds to 25 MPa. **c**, Lithostatic pressure at sea level calculated along dyke segments 1, 2,

3, 4, 5, 6, 7a, 8a and 8b (blue line). The calculations take into account both the subglacial bedrock topography and ice thickness. Light blue triangles indicate the beginning of a segment and red triangles the end of a segment. It is assumed that between segments the dyke propagates along a straight path. Dyke propagation was halted for the longest time at the end of segment 4 (see Figs 2 and 4), before an increase in lithostatic pressure.

Copulation in antiarch placoderms and the origin of gnathostome internal fertilization

John A. Long^{1,2,3}, Elga Mark-Kurik⁴, Zerina Johanson⁵, Michael S. Y. Lee^{6,7}, Gavin C. Young⁸, Zhu Min⁹, Per E. Ahlberg¹⁰, Michael Newman¹¹, Roger Jones¹², Jan den Blaauwen¹³, Brian Choo¹ & Kate Trinajstić^{14,15}

Reproduction in jawed vertebrates (gnathostomes) involves either external or internal fertilization¹. It is commonly argued that internal fertilization can evolve from external, but not the reverse. Male copulatory claspers are present in certain placoderms^{2–4}, fossil jawed vertebrates retrieved as a paraphyletic segment of the gnathostome stem group in recent studies^{5–8}. This suggests that internal fertilization could be primitive for gnathostomes, but such a conclusion depends on demonstrating that copulation was not just a specialized feature of certain placoderm subgroups. The reproductive biology of antiarchs, consistently identified as the least crownward placoderms^{5–8} and thus of great interest in this context, has until now remained unknown. Here we show that certain antiarchs possessed dermal claspers in the males, while females bore paired dermal plates inferred to have facilitated copulation. These structures are not associated with pelvic fins. The clasper morphology resembles that of ptyctodonts, a more crownward placoderm group^{7,8}, suggesting that all placoderm claspers are homologous and that internal fertilization characterized all placoderms. This implies that external fertilization and spawning, which characterize most extant aquatic gnathostomes, must be derived from internal fertilization, even though this transformation has been thought implausible. Alternatively, the substantial morphological evidence for placoderm paraphyly must be rejected.

Among living fish (non-tetrapod vertebrates) internal fertilization is mostly effected by copulatory structures representing modifications of the pelvic fins (chondrichthyans⁹) or anal fins (the gonopodium of some teleost fishes^{10,11}). The discovery in arthrodiran³ and ptyctodont placoderms^{2,4} of copulatory structures (claspers) originally^{3,12} but incorrectly⁴ interpreted as modified pelvic fins implied that any placoderms lacking pelvic appendages would have also lacked claspers, and therefore reproduced externally by spawning. This was thought to be the case with the antiarchs, recently considered the sister group of all other jawed vertebrates^{5–8}. Deposits where small antiarch juveniles suffered mass mortality have been interpreted as nurseries associated with mass spawning events¹³.

Microbrachius, a small bothriolepidoid antiarch, is known from complete articulated specimens from the Middle Devonian (Givetian) period of Scotland¹⁴, the Early–Middle Devonian of China^{15,16} and by isolated plates from the Essi Farm site in Estonia (Extended Data Figs 1 and 2 provide geological background for sites of specimens described herein). Numerous new articulated specimens from the Eday Flags, Orkney Islands, Scotland, UK, show either male dermal claspers or female genital plates in life position (Figs 1 and 2 and Extended Data Fig. 3), and one specimen from Estonia shows an isolated right posterior ventrolateral (PVL) plate with an attached dermal clasper (Fig. 1a–c) (Supplementary Information A1 and A2 and Extended Data Figs 1 and 2).

The clasper of *Microbrachius* is a deeply grooved dermal bone (Fig. 1a–c, h–m) that curves laterally, similar to the main hooked dermal clasper element of ptyctodontid placoderms^{2,4}. The groove (gr, Fig. 1l, m) may have served to transfer sperm, or encased a structure that carried the sperm canal. The articulated specimens show some of the claspers in mesial contact, sutured in the midline (Fig. 1i, k) indicating they were not mobile. The extended wing of the clasper extends laterally as far as the width of the trunkshield (Fig. 1h, j, k), potentially enabling a male *Microbrachius* clasper to reach the cloaca of a female if the two individuals were side by side (Fig. 3). Ventrally the claspers have well defined dermal ornamentation of small posteriorly directed spines, and have a series of larger spines along the distal margin (Fig. 1i, m). Variation in clasper size probably reflects individual sexual maturity (Extended Data Fig. 3).

Female *Microbrachius dicki* (Figs 2e–g and 3a, b) show paired blade-like structures in the same region corresponding to the male clasper (Fig. 2g). These blades carry a distinctive ornament of curving ridges and marginal tubercles on their dorsal (that is, internal) surfaces: the only internally facing ornament in their dermal skeleton. Identical internal ornament is seen in *Pterichthyodes*¹⁴, where it occurs on separate dermal plates preserved atop the dorsal (internal) side of the subanal lamina of the PVL plates (Fig. 2a–d) in similar position as the male claspers. Within other placoderms, where separate male and female dermal elements are known, they are similarly positioned^{2,4}. These dermal plates are flat and taper to meet the lateral ends of the transverse ventral ridge inside the lateral lamina of the PVL plate (Fig. 2f–h). We interpret these structures in *Microbrachius* and *Pterichthyodes* as female genital plates, similar to the post-pelvic plates found in some female ptyctodonts⁴, and suggest that the claspers attached to them during mating by gripping the internal ornament which faced into the cloacal chamber (Fig. 3e and Extended data Fig. 9).

Two specimens of the Late Devonian antiarch *Bothriolepis* also show small semicircular plates sitting on the distal area of the subanal lamina of the PVL plates (Fig. 2h–j). They are only visible in dorsal view, and show a slight thickening anteriorly with a line of roughened pits for ligamentous or muscle attachment near the anterior margin. We interpret these as female genital plates as they do not resemble male claspers in *Microbrachius*, and occupy the same topology as the paired genital plates in female *Pterichthyodes*. We also identify new features on the subanal lamina of the PVL plates in other antiarchs, which probably relate to their reproductive anatomy (Extended Data Fig. 4 and Supplementary Information B4).

We thus have compelling evidence for a clasper-based system of internal fertilization in *Microbrachius*, strong circumstantial evidence for the same system in *Pterichthyodes* and plausible evidence for the

¹School of Biological Sciences, Flinders University, 2100, Adelaide, South Australia 5001, Australia. ²Natural History Museum of Los Angeles County, 900 Exposition Boulevard, Los Angeles, California 9007, USA. ³Museum Victoria, PO Box 666, Melbourne, Victoria 3001, Australia. ⁴Institute of Geology at Tallinn University of Technology, Ehitajate tee 5, 19086 Tallinn, Estonia. ⁵Department of Earth Sciences, Natural History Museum, London SW7 5BD, UK. ⁶South Australian Museum, North Terrace, Adelaide, South Australia 5000, Australia. ⁷School of Earth and Environmental Sciences, The University of Adelaide, South Australia 5005, Australia. ⁸Research School of Earth Sciences, The Australian National University, Canberra, Australian Capital Territory 0200, Australia. ⁹Key Laboratory of Evolutionary Systematics of Vertebrates, Institute of Vertebrate Paleontology and Paleoanthropology, Chinese Academy of Sciences, PO Box 643, Beijing 100044, China. ¹⁰Department of Organismal Biology, Evolutionary Biology Centre, Uppsala University, Norbyvägen 18A, 752 36 Uppsala, Sweden. ¹¹Vine Lodge, Vine Road, Johnston, Haverfordwest, Pembrokeshire SA62 3NZ, UK. ¹²6 Burghley Road, Wimbledon, London SW19 5BH, UK. ¹³University of Amsterdam, Science Park 904, 1098XH, Amsterdam, The Netherlands. ¹⁴Western Australian Organic and Isotope Geochemistry Centre, Department of Chemistry, Curtin University, Perth, Western Australia 6102, Australia. ¹⁵Earth and Planetary Sciences, Western Australian Museum, Perth, Western Australia 6000, Australia.

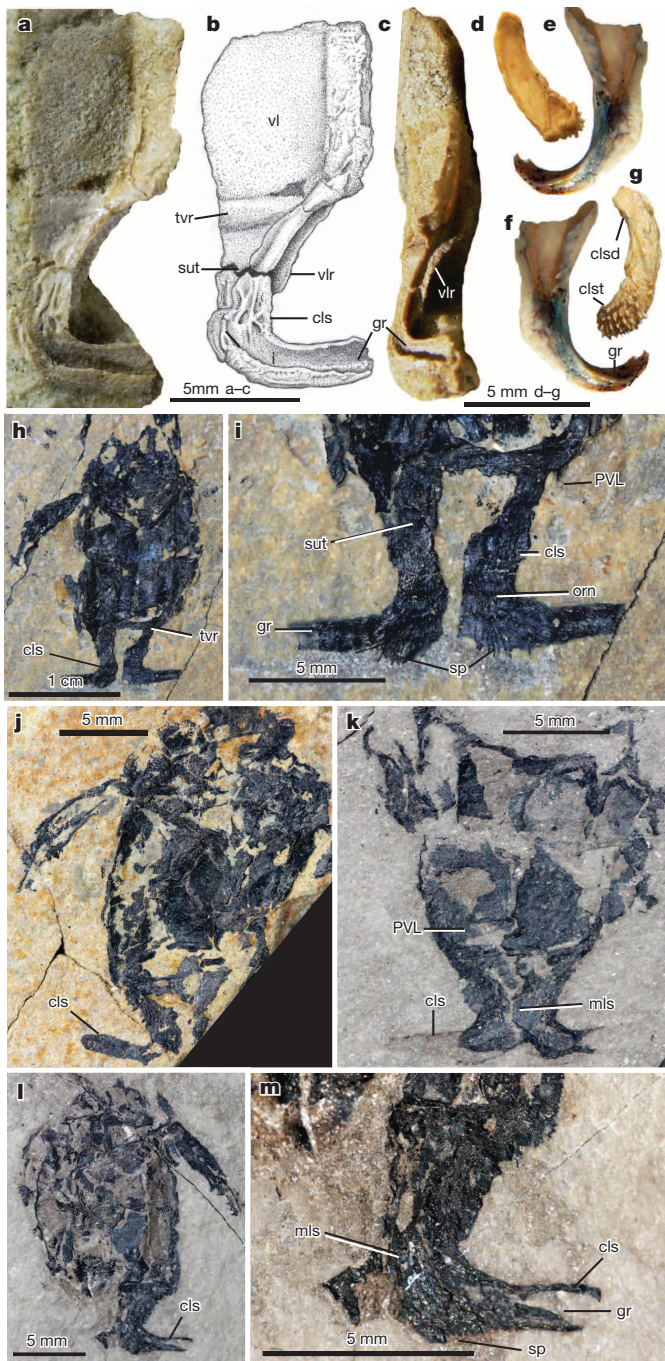


Figure 1 | Male reproductive structures in antiarchs and ptyctodontids. **a–c**, *Microbrachius* sp. right PVL plate, GIT 628-24 in **(a)** dorsal view, **(b)** interpretive drawing and **(c)** lateral view. **d–g**, Dermal clasp elements of ptyctodont *Austroptyctodus gardineri*, NHMUK PV P57665; **d**, dermal hooked elements in dorsal and ventral **(g)** views; **e**, large hooked dermal element in ventral view; **f**, reversed image. **h–m**, Male *M. dicki* specimens showing claspers with close-up views of some specimens. **h**, **i**, NHMUK PV P73397; **j**, NHMUK PV P77402; **k**, NHMUK PV P77401; **l**, **m**, NHMUK PV P77405. Abbreviations: cls, dermal clasp elements; clsd, distal clasp element; clst, terminal hooked clasp element; gr, groove; mls, midline suture; orn, ornament; PVL, posterior ventrolateral plate; ri, ridge; sp, spines; sut, suture between clasper and PVL; tvr, transverse lateral ridge; vl, ventral lamina; vlr, ventrolateral ridge.

same in *Bothriolepis*. Although male claspers have not been described in antiarchs other than *Microbrachius*, even from taxa such as *Bothriolepis*^{17,18}, *Asterolepis*¹⁹ and *Remigolepis*²⁰ known from hundreds of articulated specimens with tails preserved, we suggest that internal fertilization

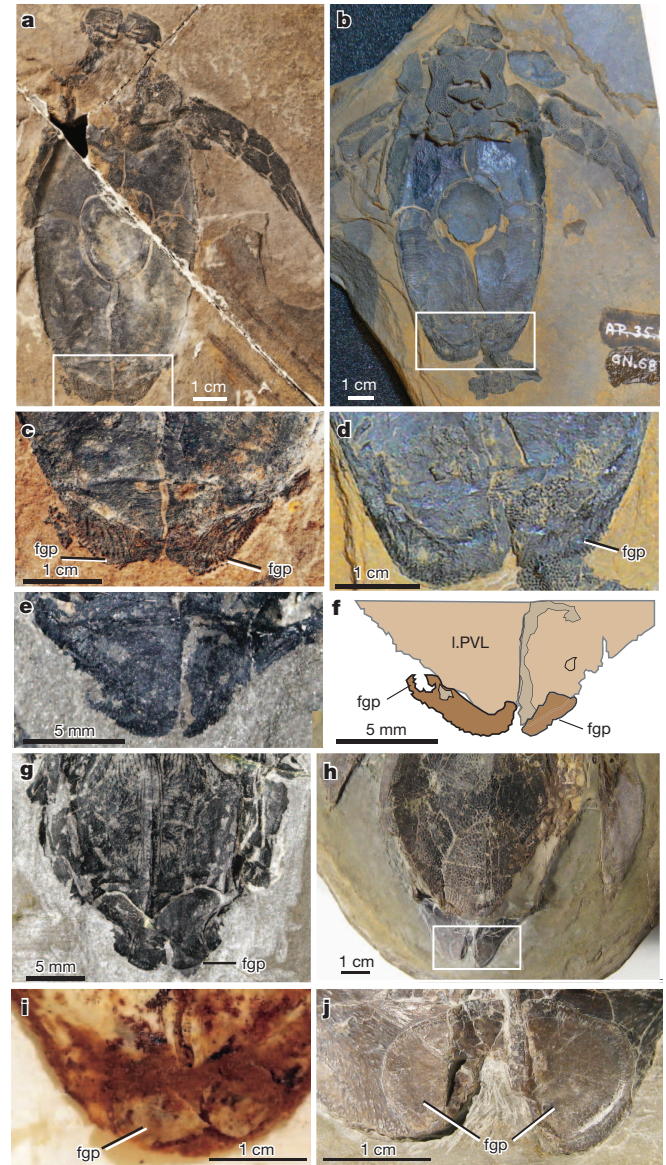


Figure 2 | Female reproductive structures in Middle–Late Devonian antiarchs. **a, c**, *Pterichthyodes milleri*; NHMUK PV P32544; **b, d**, UMZC 687. **e–g**, *M. dicki*; **e, f**, NHMUK PV P73398; **g**, NHMUK PV P73399. **h, i**, *Bothriolepis canadensis*, V11127 (Smithsonian), showing armour (**h**) and female genital plates (**i**). **j**, *Bothriolepis* sp. ANU V1040, close up of paired female genital plates. Abbreviations: fgp, female genital plates; LPVL, left posterior ventrolateral plate.

is general for the Antiarchi. As the dermal skeleton is quite reduced in advanced antiarchs, we propose that the claspers in some forms might have also been cartilaginous and thus not well preserved. Further evidence for internal fertilization in antiarchs comes from their large hatchlings¹³ (Supplementary Information B4).

The clasper in *Microbrachius* is clearly different from the pelvic girdle and fin, known only in one antiarch: *Parayunnanolepis* from the Early Devonian of China²¹. More derived asterolepidoid and bothriolepidoid antiarchs lack pelvic girdles and fins, which are thus assumed to be lost (secondarily absent) in these groups¹⁹. In chondrichthyans, the clasper is attached to the posterior extremity of the pelvic fin metapterygium^{9,22,23}. In ptyctodont and arthrodire placoderms, the clasper is immediately posterior to the pelvic fin⁴, and has previously been interpreted as an elaboration of the pelvic fin skeleton^{2,3}. However, new evidence shows that the clasper of arthrodires does not articulate directly with the pelvic girdle or fin⁴. In ptyctodonts the endoskeleton of the clasper (if present)

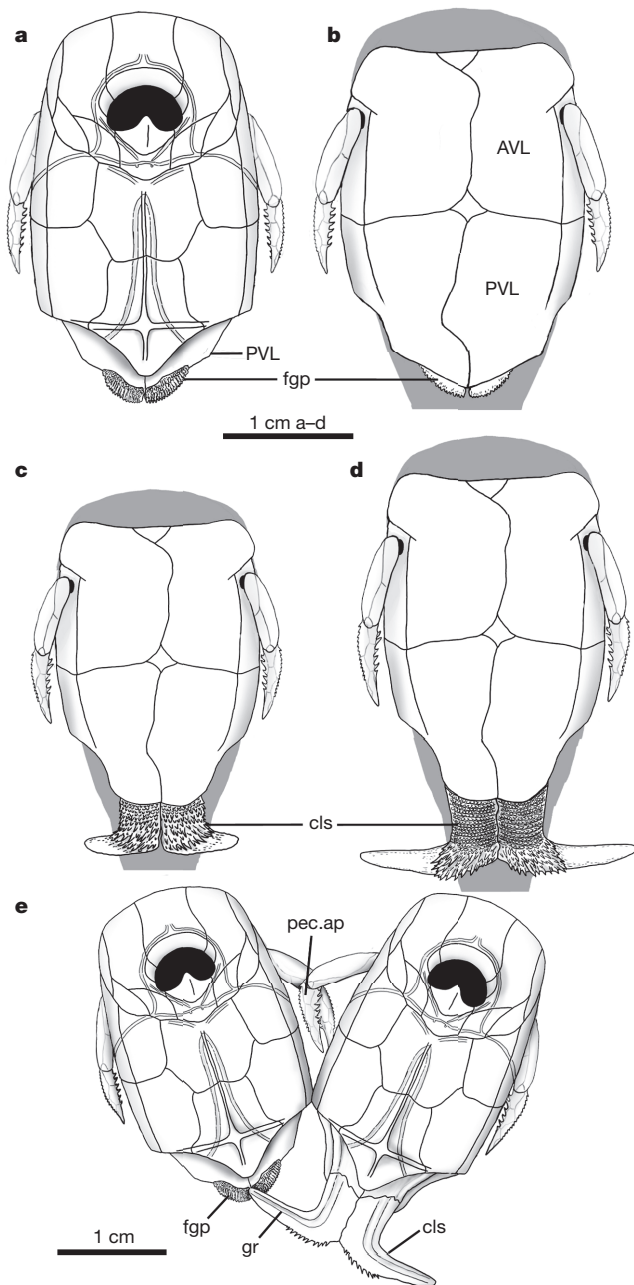


Figure 3 | Male and female sexual dimorphism in *M. dicki*. Reconstruction of female *M. dicki* in (a) dorsal and (b) ventral views; c, d, male *M. dicki* ventral views showing variations in clasper development; e, hypothetical mating position for *Microbrachius*. Abbreviations as for Figs 1 and 2 plus pec.ap, pectoral appendage.

was unossified¹², making it more difficult to determine its precise relationship to neighbouring structures, but the dermal components of the claspers are consistently preserved some distance posterior to the pelvis^{2,4,12}. All known placoderm claspers thus differ from chondrichthyan claspers in being independent from the pelvis and pelvic fin. The principle dermal element of the ptyctodont clasper is a large curved, grooved plate (Fig. 1e, f) that distinctly resembles and is thus probably homologous to the dermal clasper of *Microbrachius*, whereas arthrodire claspers have a much smaller but plausibly homologous dermal bone tip^{3,4}. Based on similarities in position (behind and independent of pelvic region), and materials (dermal bone), the claspers of antiarchs, ptyctodonts and arthrodires are most probably homologous with each other, but not homologous with the claspers of chondrichthyans.

To evaluate the evolution of claspers and reproductive biology across gnathostomes, we expanded upon a recently published phylogenetic analysis⁸ with the addition of 14 placoderm taxa, three new characters (256–258), and one character (122) split into two (122, 159) (Supplementary Information C7, 8). Our analyses of both expanded and original data sets yielded very similar trees, which supported placoderm paraphyly and placed antiarchs as the sister group to all other gnathostomes (Extended Data Figs 5 and 6 and Supplementary Information C7, 8). Our analyses found shorter trees for the original data set⁸, supporting a more orthodox position for ptyctodonts lower on the gnathostome stem (Extended Data Figs 7 and 8 and Supplementary Information C8). The shortest tree indicates that bony claspers separate from the pelvic fin arose in the most recent common ancestor of jawed vertebrates and were lost in the most recent common ancestor of crown gnathostomes (Fig. 4 and Extended Data Figs 5 and 6).

These results have intriguing biological implications. The implied loss of bony claspers and implied reversion to external fertilization in crown gnathostomes appears heterodox: loss of internal fertilization and acquisition of external fertilization is not widely accepted, at least

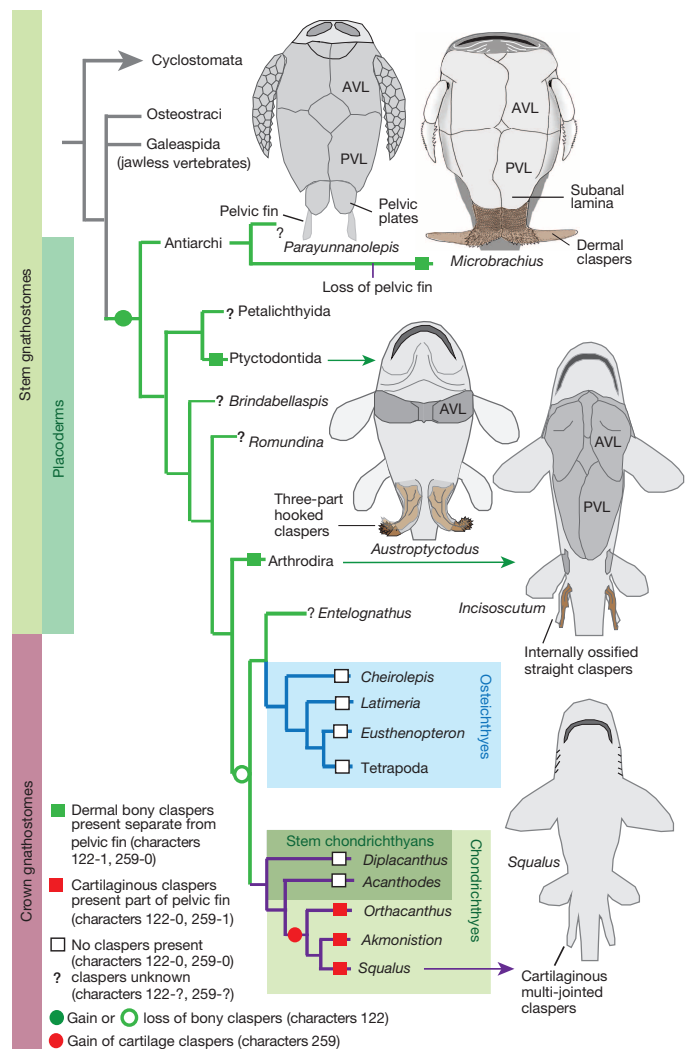


Figure 4 | Phylogeny of major lineages of gnathostomes, based on analysis of an expanded version of the data set from ref. 8. Distribution and morphology of two kinds of clasper are mapped on the phylogeny. Claspers are most parsimoniously inferred to have evolved in the most recent common ancestor of all gnathostomes, then lost before or at the node of crown group gnathostomes. Claspers developed as a modified part of the pelvic fin skeleton appear as a synapomorphy of all chondrichthyans. The full tree with branch supports and states for all terminal taxa is shown in Extended Data Figs 5 and 6. See Supplementary Information for further details of the analysis.

in vertebrates^{1,11} although it could have happened multiple times in invertebrates²⁴. The shared, unique morphology and post-pelvic position of claspers in all placoderms is more consistent with a single origin, and thus represents a potential synapomorphy supporting placoderm monophyly²⁵.

If all placoderm claspers are homologous, as we suggest, this gives rise to alternative implications with equally startling significance for early vertebrate evolution. If placoderm paraphyly is accepted, on the basis of the optimal trees here (and consistent with the majority of recent analyses^{5–8}), then external fertilization and spawning employed by the majority of recent bony fishes and many lissamphibians must have evolved from clasper-mediated internal fertilization. If claspers are accepted as *prima facie* evidence of placoderm monophyly²⁵, the transformation of cranial architecture within the Placodermi documented by several recent analyses^{5–8} must be entirely convergent on crown gnathostomes, as the antiarchs belong to the cluster of primitive placoderms with posteriorly placed rostronasal capsules and a trabecular ‘upper lip’⁸. Placoderm monophyly is also inconsistent with the co-occurrence of placoderm and osteichthyan characteristics in Silurian taxa such as *Entelognathus*⁷. Resolution of the status of placoderms, which will require data both from fossil anatomy and the reproductive physiology of extant fishes, is one of the most pressing tasks currently facing early gnathostome palaeontology.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 27 January; accepted 29 August 2014.

Published online 19 October 2014.

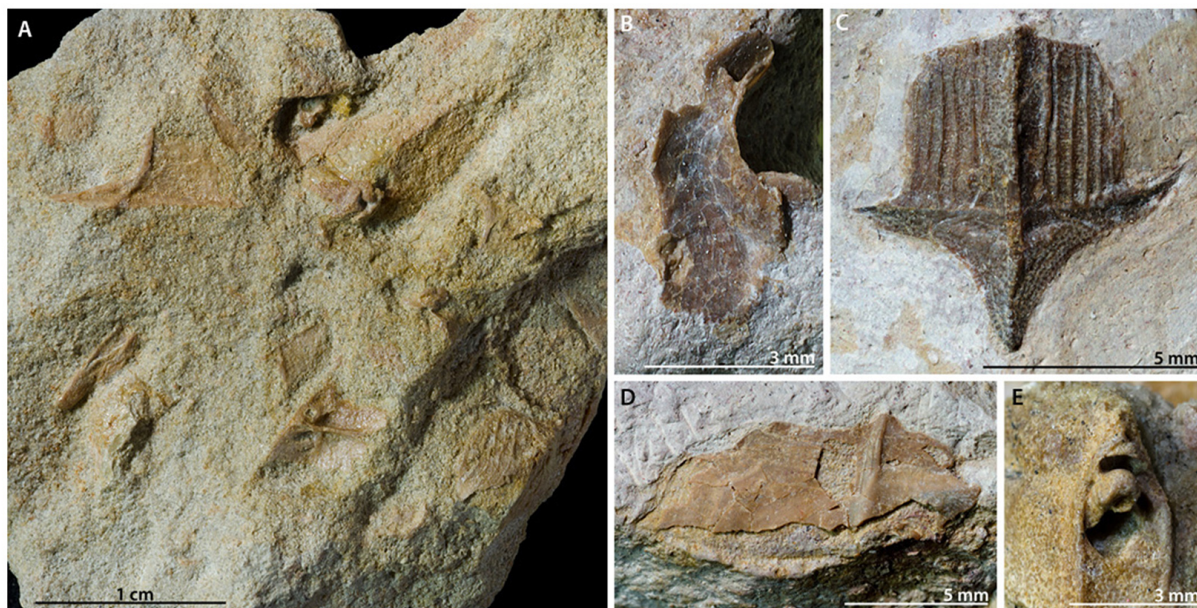
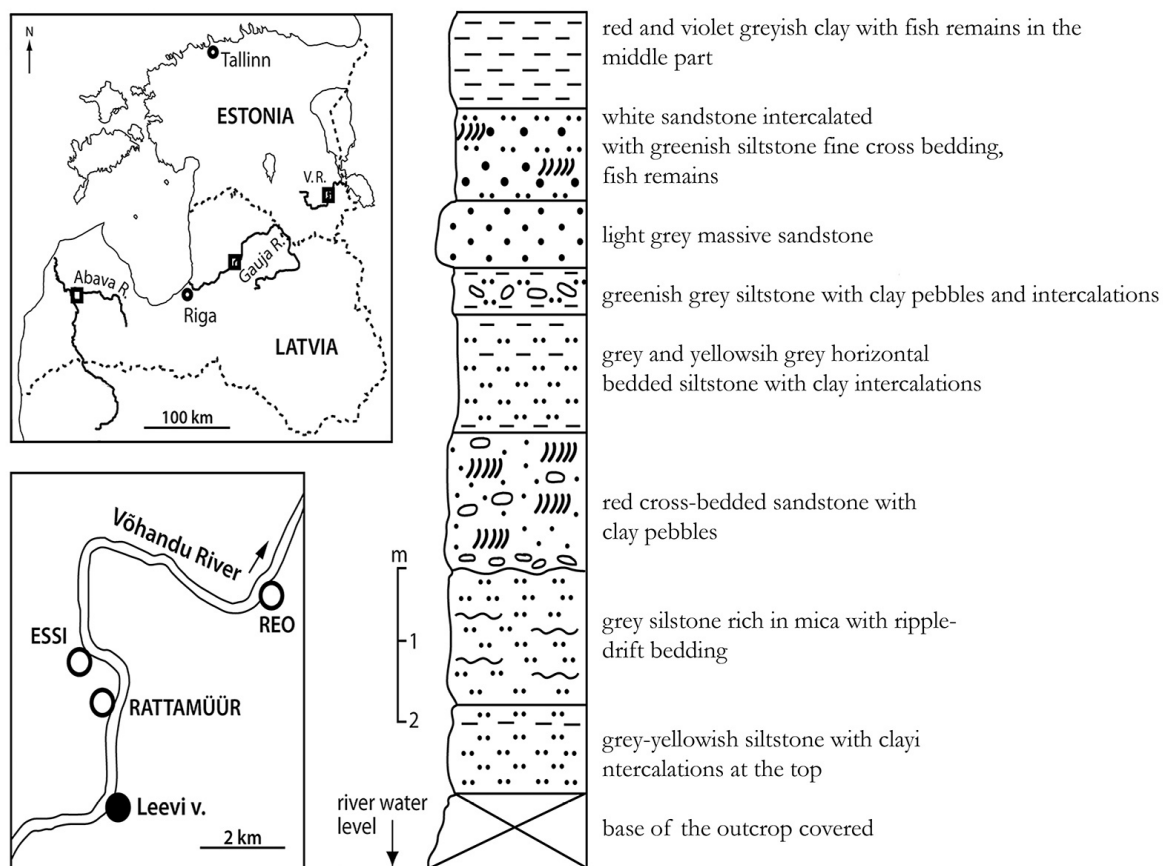
- Blackburn, D. G. Evolution of vertebrate viviparity and specializations for fetal nutrition: a quantitative and qualitative analysis. *J. Morphol.* <http://dx.doi.org/10.1002/jmor.20272> (2014).
- Miles, R. S. Observations on the ptyctodont fish, *Rhamphodopsis* Watson. *Zool. J. Linn. Soc.* **47**, 99–120 (1967).
- Ahlberg, P. E., Trinajstić, K., Johanson, Z. & Long, J. A. Pelvic claspers confirm chondrichthyan-like internal fertilization in arthrodires. *Nature* **460**, 888–889 (2009).
- Trinajstić, K. M., Boisvert, C., Long, J. A., Maksimeko, A. & Johanson, Z. Pelvic and reproductive structures in placoderms (stem gnathostomes). *Biol. Rev.* <http://dx.doi.org/10.1111/brv.12118> (2014).
- Brazeau, M. D. The braincase and jaws of a Devonian ‘acanthodian’ and modern gnathostome origins. *Nature* **457**, 305–308 (2009).
- Davis, S. P., Finarelli, J. A. & Coates, M. I. *Acanthodes* and shark-like conditions in the last common ancestor of modern gnathostomes. *Nature* **486**, 247–250 (2012).
- Zhu, M. *et al.* A Silurian placoderm with osteichthyan-like marginal jaw bones. *Nature* **502**, 188–193 (2013).
- Dupret, V., Sanchez, S., Goujet, D., Tafforeau, P. & Ahlberg, P. E. A primitive placoderm sheds light on the origin of the jawed vertebrate face. *Nature* **507**, 500–503 (2014).
- Freitas, R., Zhang, G. & Cohn, M. J. Biphase *Hoxd* gene expression in shark paired fins reveals an ancient origin of the distal limb domain. *PLoS ONE* **2**, e754 (2007).
- Meyer, A. & Lydeard, C. The evolution of copulatory organs, internal fertilization placenta and viviparity in killifishes (Cyprinodontiformes) inferred from a DNA phylogeny of the tyrosine kinase gene *X-src*. *Proc. R. Soc. Lond. B* **254**, 153–162 (1993).
- Parenti, L. R., LoNostro, F. L. & Grier, H. J. Reproductive histology of *Tomeurus gracilis* Eigenmann, 1909 (Teleostei: Atherinomorpho: Poeciliidae) with comments on evolution of viviparity in atherinomorph fishes. *J. Morphol.* **271**, 1399–1406 (2010).
- Miles, R. S. & Young, G. C. Placoderm interrelationships reconsidered in the light of new ptyctodontids from Gogo, Western Australia. *Linn. Soc. Symp. Ser.* **4**, 123–198 (1977).
- Downs, J. P., Criswell, K. E. & Daeschler, E. B. Mass mortality of juvenile antiarchs (*Bothriolepis* sp.) from the Catskill Formation (Upper Devonian, Famennian Stage), Tioga county, Pennsylvania. *Proc. Acad. Nat. Sci. Philad.* **161**, 191–203 (2011).
- Hemmings, S. K. The Old Red Sandstone antiarchs of Scotland: *Pterichthyodes* and *Microbrachius*. *Palaeontogr. Soc. Monogr.* **131** (551), 1–64 (1978).
- Pan, J. A new species of *Microbrachius* from the Middle Devonian of Yunnan. *Vertebr. Palasiat.* **22**, 8–13 (1984).
- Wang, J.-Q. & Zhang, G.-R. New material of *Microbrachius* from the Lower Devonian of Qujing, Yunnan, China. *Vertebr. Palasiat.* **37**, 200–211 (1999).
- Stensiö, E. A. On the Placodermi from the Upper Devonian of East Greenland. II. Antiarchi: subfamily Bothriolepinae. With an attempt at a revision of the previously described species of the family. *Medd. Grönl.* **139**, 1–622 (1948).
- Long, J. A. & Werdelin, L. A new Late Devonian bothriolepid (Placodermi, Antiarcha) from Victoria, with descriptions of others from the state. *Alcheringa* **10**, 355–399 (1986).
- Lysarkaya, L. A. *Baltic Devonian Placodermi. Asterolepididae* [in Russian] 1–152 (Zinatne, 1981).
- Johanson, Z. New *Remigolepis* (Placodermi; Antiarchi) from Canowindra, New South Wales, Australia. *Geol. Mag.* **134**, 813–846 (1997).
- Zhu, M., Yu, X., Choo, B., Wang, J. & Jia, L. An antiarch placoderm shows that pelvic girdles arose at the root of jawed vertebrates. *Biol. Lett.* **8**, 453–456 (2012).
- Leigh-Sharpe, H. The comparative morphology of the secondary sexual characters of elasmobranch fishes. The claspers, clasper siphons, and clasper glands. Memoir III–V. *J. Morphol.* **36**, 190–240 (1922).
- Goodrich, E. S. *Studies on the Structure and Development of Vertebrates* Vol. 1, 1–486 (Dover Publications, 1958).
- Rouse, G. W. Broadcasting fables: is external fertilization really primitive? Sex, size, and larvae in sabellid polychaetes. *Zool. Scr.* **23**, 271–312 (1994).
- Brazeau, M. D. & Friedman, M. The characters of Palaeozoic jawed vertebrates. *Zool. J. Linn. Soc.* **170**, 779–821 (2014).

Supplementary Information is available in the online version of the paper.

Acknowledgements For access to collections we thank E. Bernard, D. Pickering, E. Fitzgerald, L. Grande, W. Simpson and H.-D. Sues. For photography we thank P. Hurst, G. Baranov and D. Hubert. We thank M. Brazeau for reviewing an earlier version of the paper. Travel to Tallinn, London and Washington for J.A.L. to examine specimens was supported by Flinders University. G.C.Y., J.A.L. and K.T. acknowledge support from the Australian Research Council.

Author Contributions The project was designed by J.A.L., with material examined and described by J.A.L., E.M.K., Z.J., K.T., B.C. and P.E.A. M.S.Y.L. performed phylogenetic analyses with input from J.A.L., G.C.Y. and B.C. M.N., J.D.B. and R.J. collected and prepared material, provided site information and input to the discussion. Illustrations were made by J.A.L. and B.C. with photography supplied for some specimens by the Natural History Museum, London, and by the Institute of Geology at Tallinn University of Technology. All authors contributed to data interpretation, figures and writing of the paper.

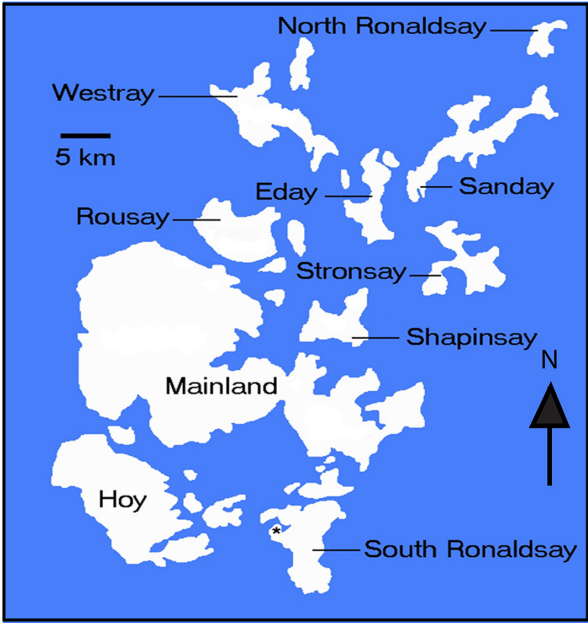
Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to J.A.L. (john.long@flinders.edu.au).



Extended Data Figure 1 | Location and stratigraphic position of Estonian specimens of *Microbrachius*. Top, map showing locality of the Essi Farm site, Estonia, and stratigraphical section where the fossils were found. Modified from ref. 26. Below, *Microbrachius* sp. plates from Essi Farm, Estonia. **a**, GIT 628-37, sample showing several small plates and fragments; **b**, GIT 628-9, right

lateral plate, visceral view; **c**, 628-3, posterior median dorsal plate, dorsal view; **d**, GIT 628-25, right posterior ventrolateral plate, visceral view; **e**, GIT 628-18, anterior section of anterior ventrolateral plate, lateral view showing brachial process. All specimens held within the Institute of Geology at Tallinn University of Technology, Estonia, collection GIT 628.

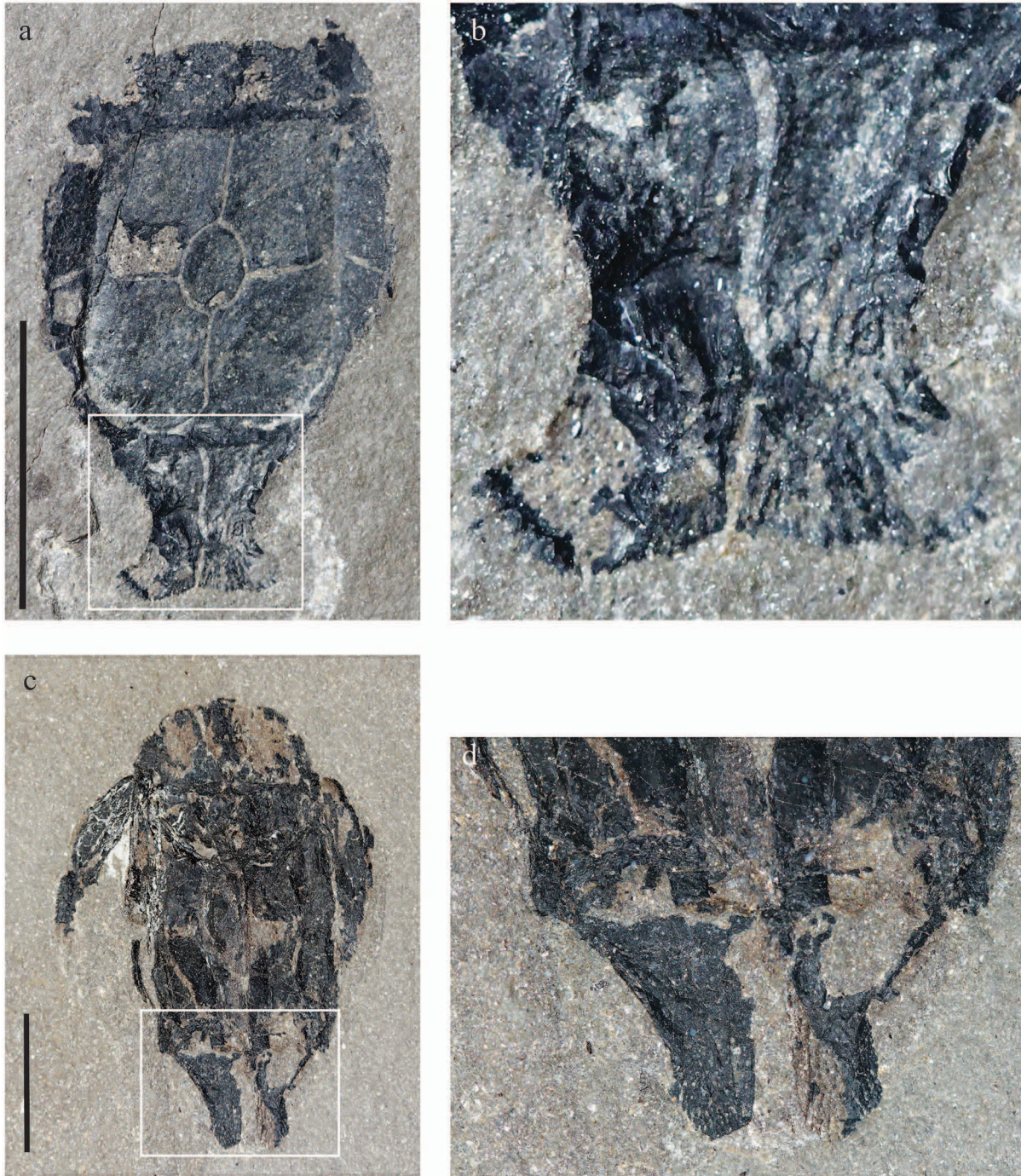
26. Mark-Kurik, E. Psammosteoid microremains from the Middle Devonian (Givetian) of Estonia. *Mod. Geol.* **24**, 1–21 (1999).



Givetian	Eday Group	Upper Eday Sandstone Formation
		Eday Marl Formation
		Middle Eday Sandstone Formation
		Eday Flagstone Formation

---?--- Eifelian	Upper Caithness Flagstone Group	Lower Eday Sandstone Formation
		Rousay Flagstone Formation

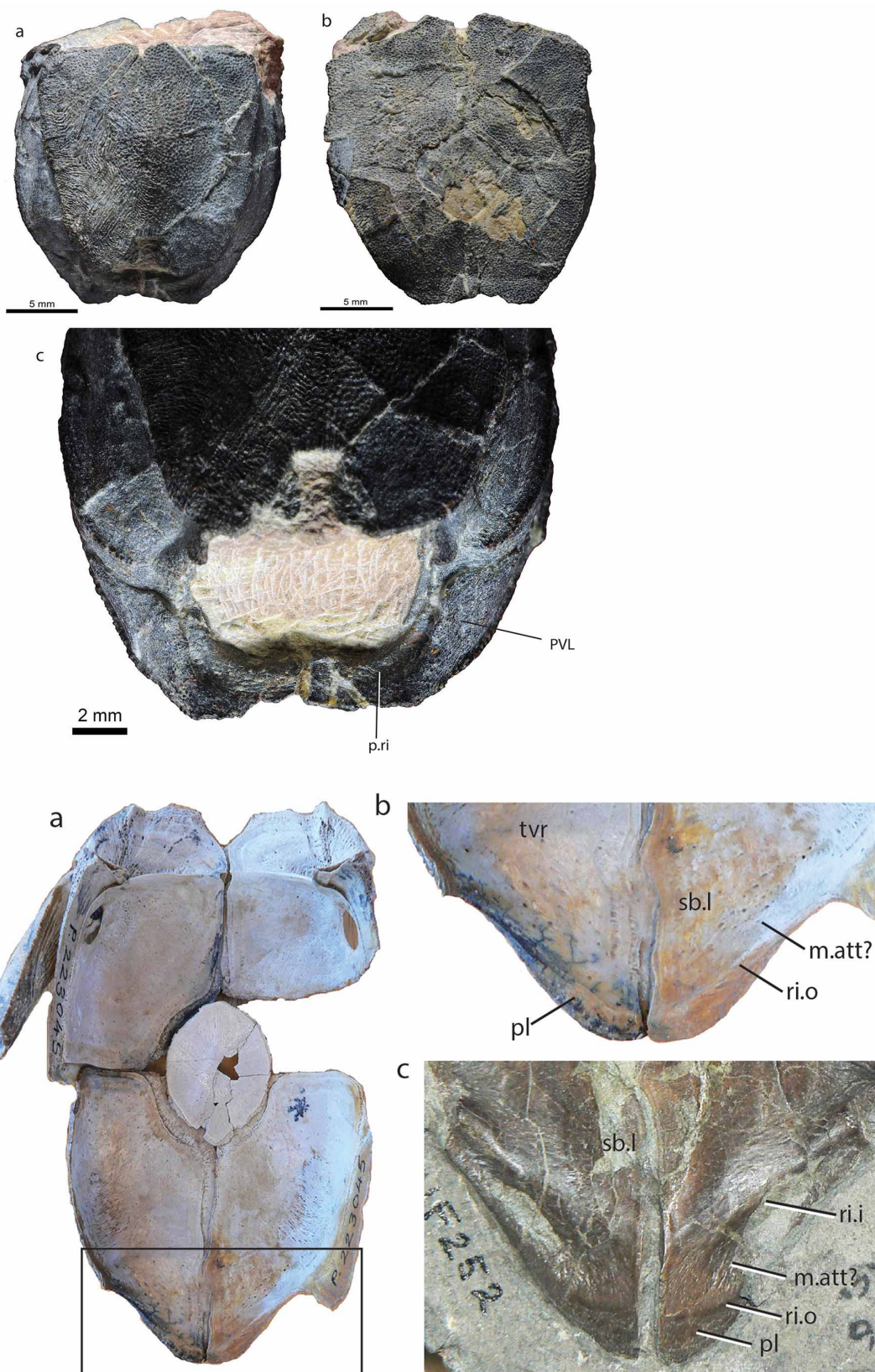
Extended Data Figure 2 | Location and stratigraphic position of new Scottish specimens of *M. dicki* described herein. Top, map of the Orkney Islands with an asterisk marking the location where the specimens of *M. dicki* described in this paper were collected. Below, stratigraphical column of the upper part of the Middle Devonian in the Orkney Islands with the position of the Eday Flagstone Formation fish beds marked by a dotted line.



Extended Data Figure 3 | Growth of claspers in *M. dicki* males.

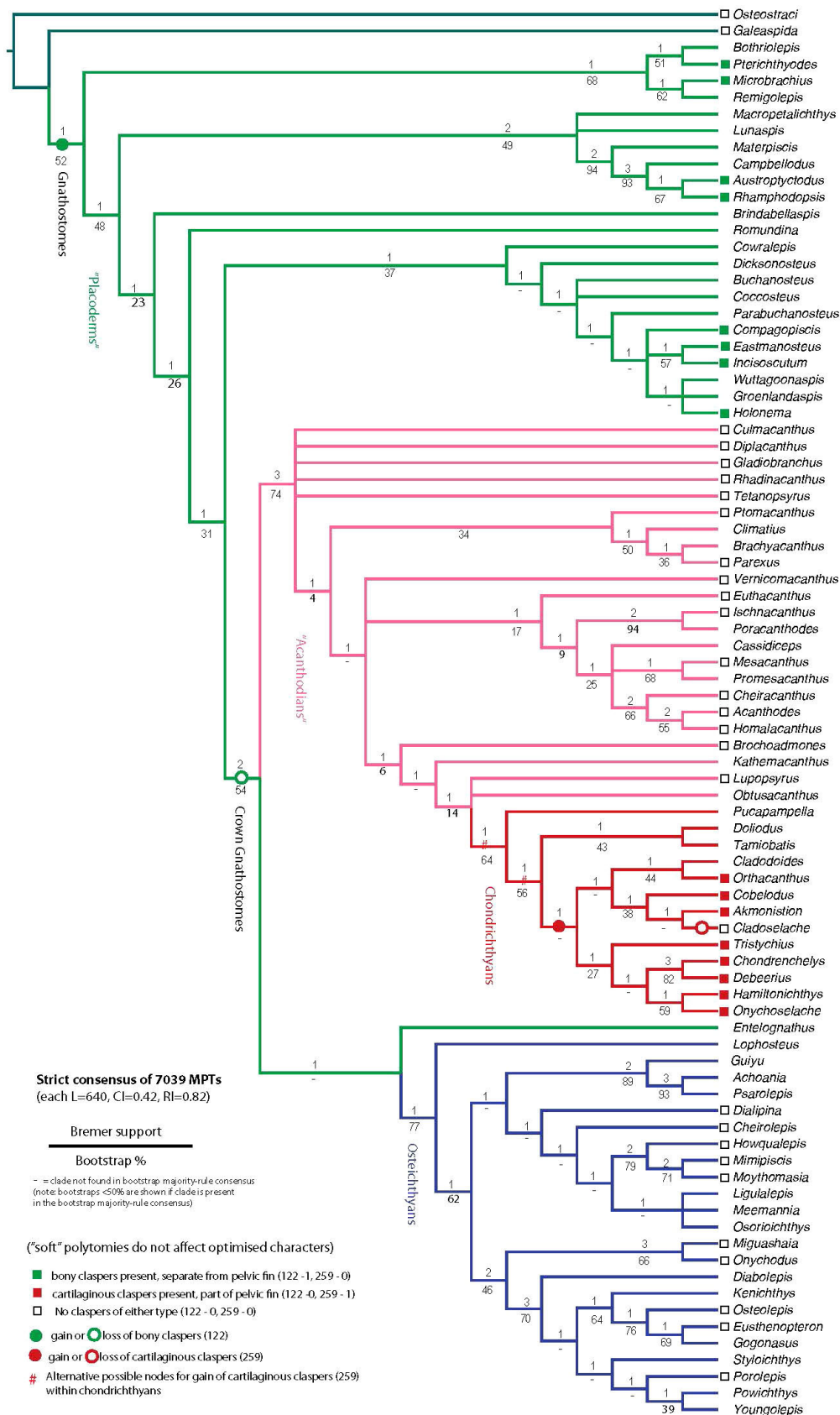
a, b, NHMUKVP P 77400, claspers only weakly developed, no lateral wing; close up of claspers in **b**; **c, d,** NHMUK VP P 77403 showing further caudally

directed growth of claspers; **d,** close up of claspers showing fusion in midline. Scale bars, 1 cm.



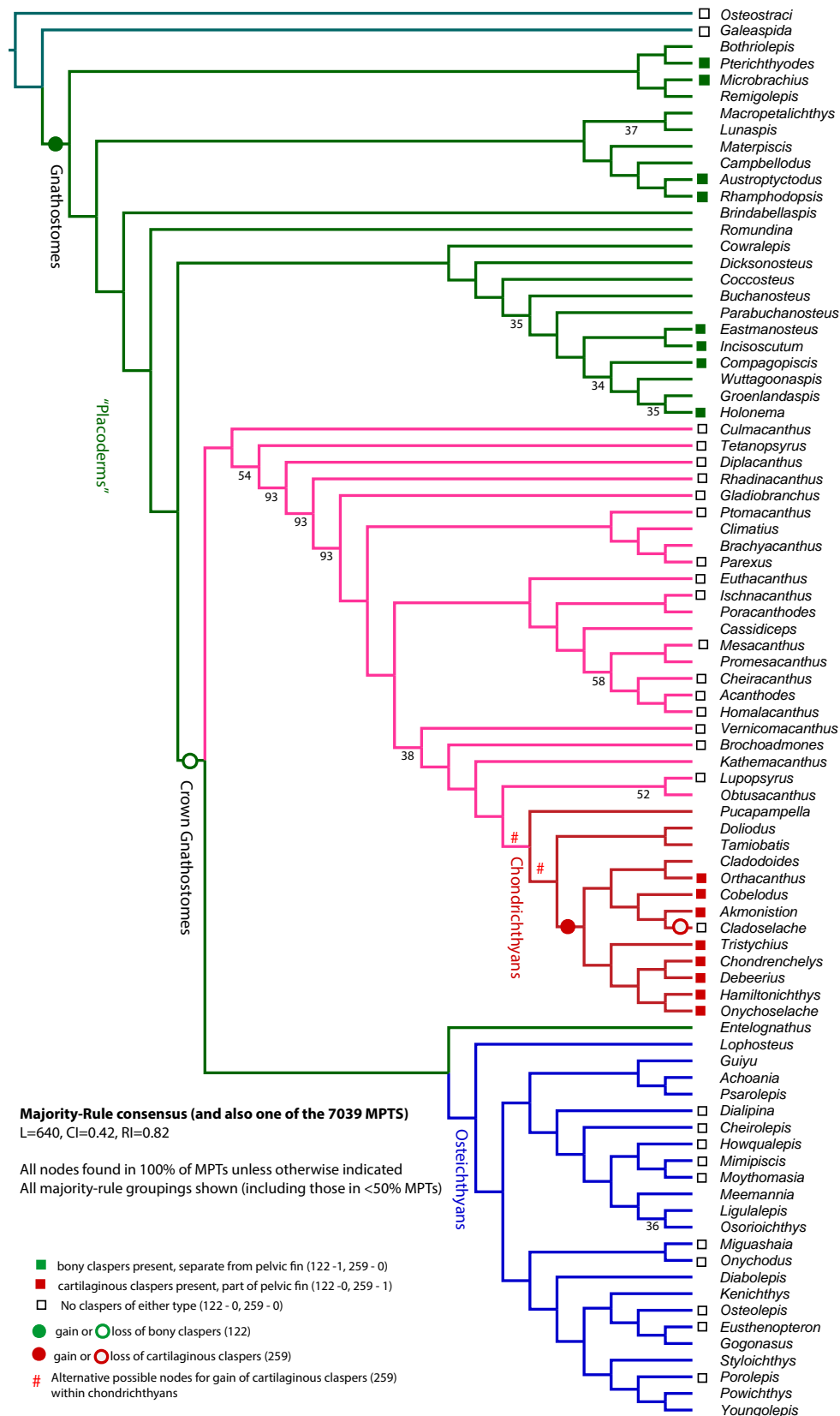
Extended Data Figure 4 | New information on pelvic region anatomy in antiarchs. Top, *Yunnanolepis porifera*, Xitun Formation, Yunnan, China. Specimen IVPP V19359 in (a) dorsal view, (b) ventral view and (c) showing posterior region of trunkshield prepared to show internal side of the PVL plates. p.ri, strong ridge on the dorsal surface of the posterior region of the PVL plates. Below, a, b, *Bothriolepis* sp., Gogo Formation, Western

Australia (P223045, Museum Victoria, Melbourne); c, *B. canadensis*, Escuminac Formation, Quebec, Canada (UF 252, Field Museum, Chicago). Abbreviations: m.att?, muscle attachment area; ri.i, internal ridge, ri.o, outer ridge; pl, platform; sb.l, subanal lamina; tvr, transverse ridge (= crista transversalis interna posterior, Stensiö 1948).



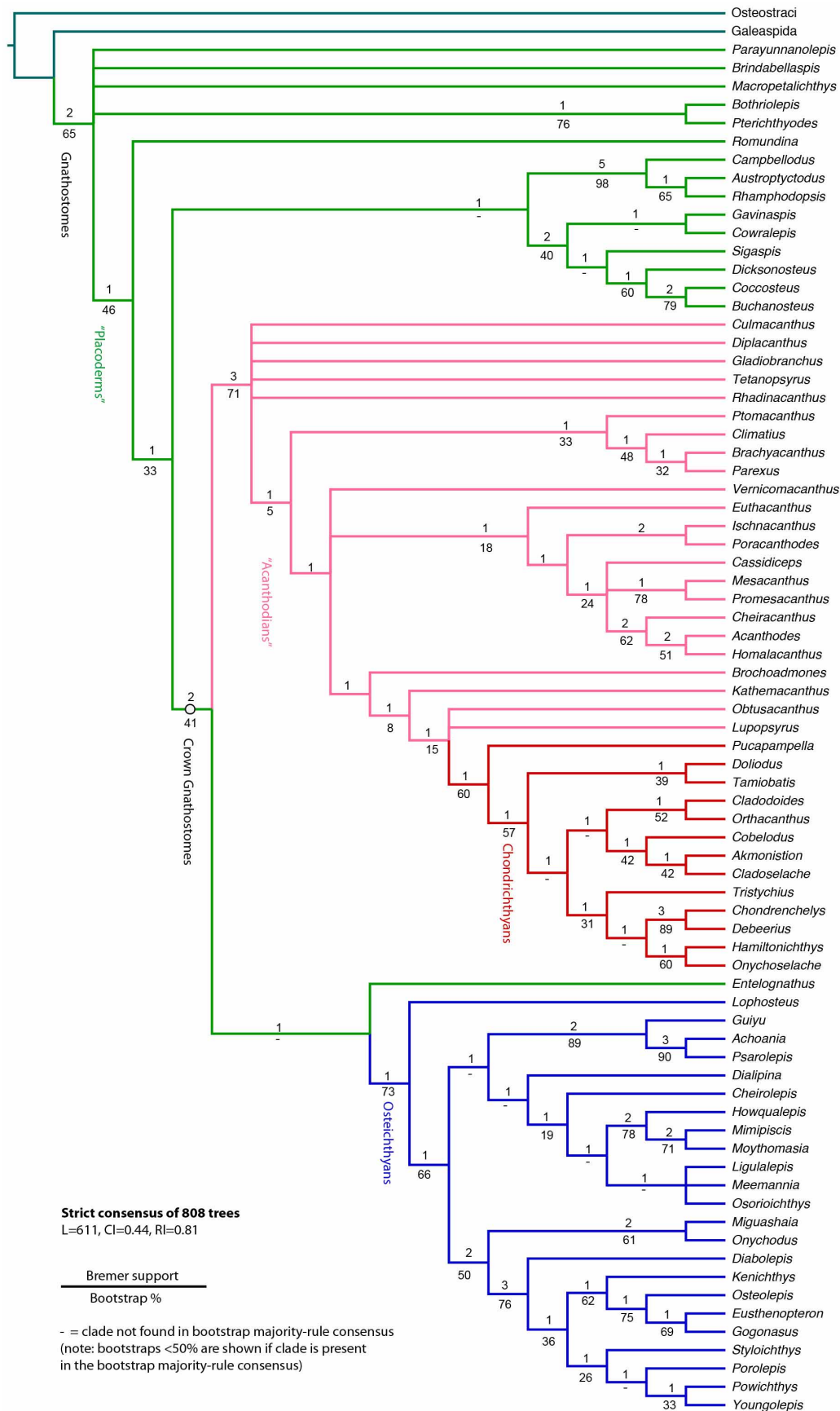
Extended Data Figure 5 | Strict consensus tree from 7,039 trees ($L = 640$) from analysis of the expanded data set (85 taxa, 259 characters). Numbers on branches denote Bremer and bootstrap support. Green squares denote presence of bony claspers (character 122), red squares denote presence of

cartilaginous claspers (character 259) and white squares denote absence of both types of clasper. Circles denote gain/loss of the two types of clasper under the most-parsimonious optimization.

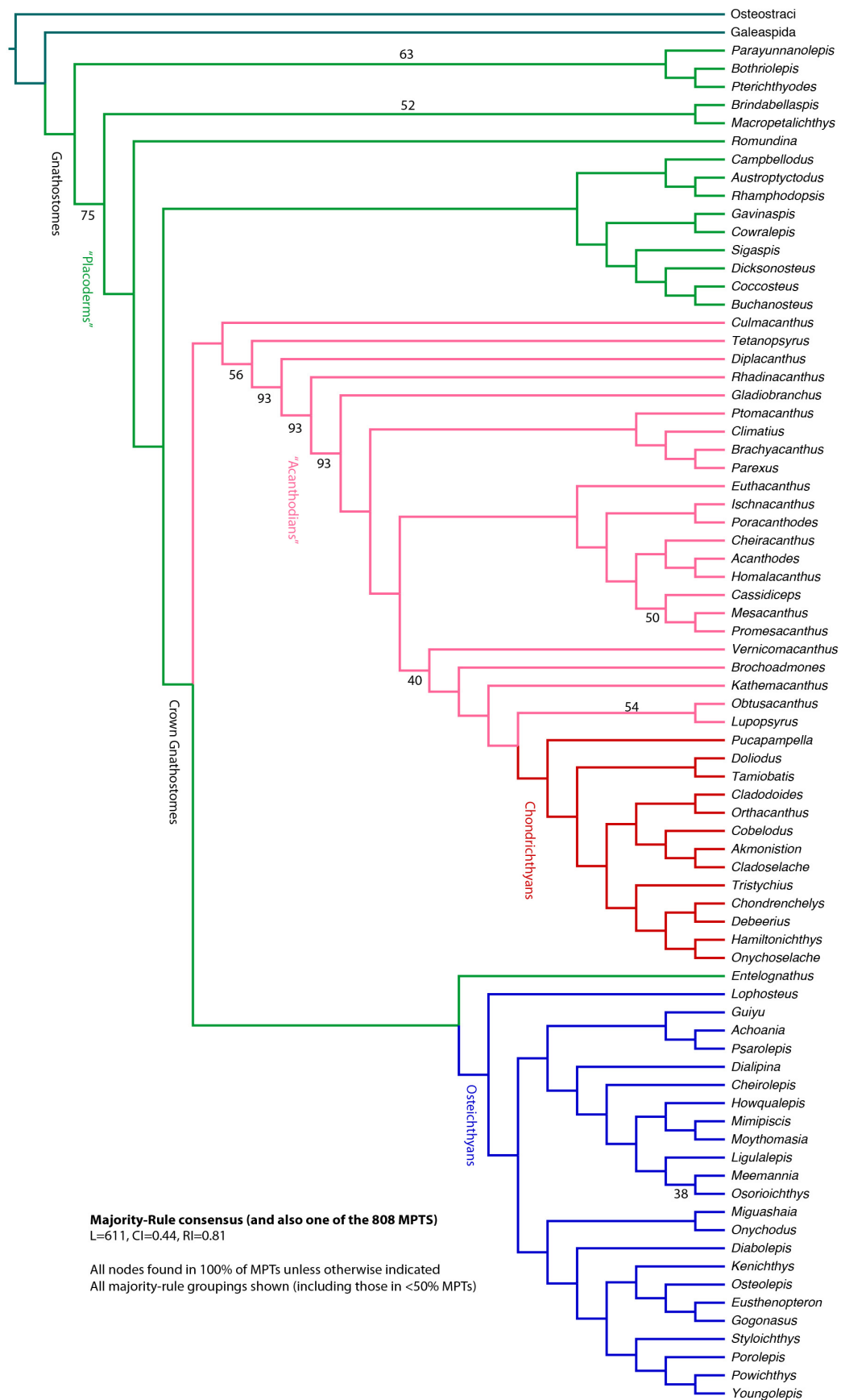


Extended Data Figure 6 | Majority-rule consensus tree, and one of the most-parsimonious trees (length 640) from analysis of the expanded data set (85 taxa, 259 characters). Numbers on branches indicate the percentage of most-parsimonious trees that contain a particular clade (100% unless otherwise

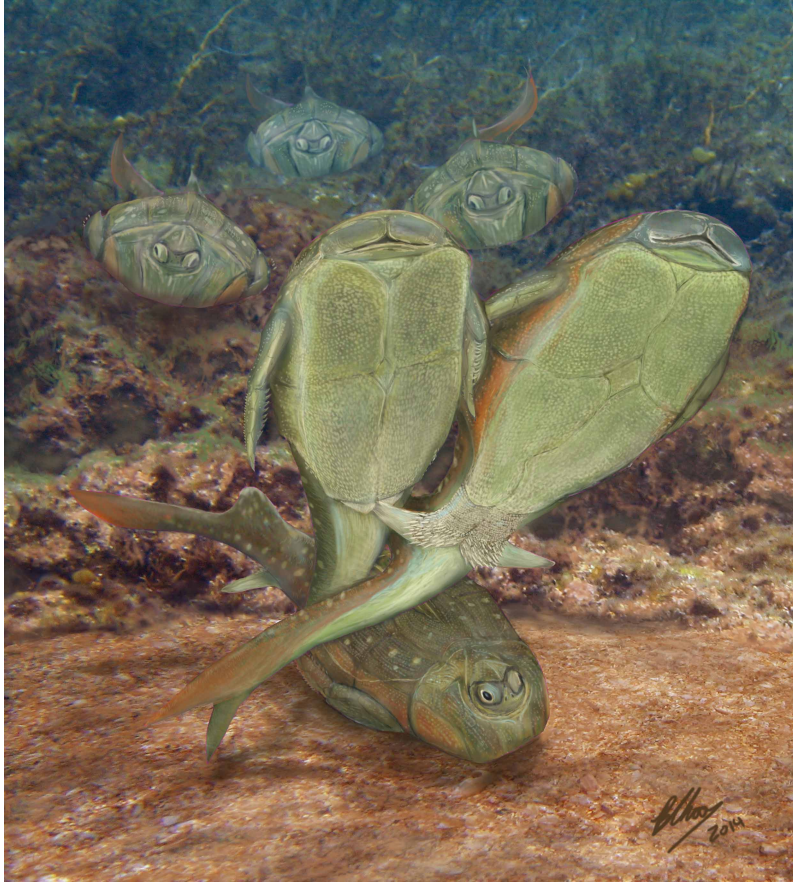
indicated). Green squares denote presence of bony claspers (character 122), red squares denote presence of cartilaginous claspers (character 259) and white squares denote absence of both types of clasper. Circles denote gain/loss of the two types of clasper under the most-parsimonious optimization.



Extended Data Figure 7 | Strict consensus tree from 808 trees ($L = 611$) from re-analysis of the data set in ref. 8. Numbers on branches denote Bremer and bootstrap support.



Extended Data Figure 8 | Majority-rule consensus tree, and one of the most-parsimonious trees (length 611) from analysis of the data in ref. 8. Numbers on branches indicate percentage of MPTs that contain a particular clade (100% unless otherwise indicated).



Extended Data Figure 9 | Reconstruction showing hypothetical mating *Microbrachius*, with male to the right, female on left. Artwork by B. Choo.

Calcium transient prevalence across the dendritic arbour predicts place field properties

Mark E. J. Sheffield¹ & Daniel A. Dombeck¹

Establishing the hippocampal cellular ensemble that represents an animal's environment involves the emergence and disappearance of place fields in specific CA1 pyramidal neurons^{1–4}, and the acquisition of different spatial firing properties across the active population⁵. While such firing flexibility and diversity have been linked to spatial memory, attention and task performance^{6,7}, the cellular and network origin of these place cell features is unknown. Basic integrate-and-fire models of place firing propose that such features result solely from varying inputs to place cells^{8,9}, but recent studies^{3,10} suggest instead that place cells themselves may play an active role through regenerative dendritic events. However, owing to the difficulty of performing functional recordings from place cell dendrites, no direct evidence of regenerative dendritic events exists, leaving any possible connection to place coding unknown. Using multi-plane two-photon calcium imaging of CA1 place cell somata, axons and dendrites in mice navigating a virtual environment, here we show that regenerative dendritic events do exist in place cells of behaving mice, and, surprisingly, their prevalence throughout the arbour is highly spatiotemporally variable. Furthermore, we show that the prevalence of such events predicts the spatial precision and persistence or disappearance of place fields. This suggests that the dynamics of spiking throughout the dendritic arbour may play a key role in forming the hippocampal representation of space.

CA1 pyramidal cell dendrites contain voltage-gated calcium and sodium channels along with NMDA (*N*-methyl-D-aspartate) receptors that allow them to produce nonlinear, regenerative (spiking) events. The spatial extent and site of generation of dendritic regenerative events can vary from widespread back-propagation of somatic action potentials (bAPs) into the arbour^{11,12} and multi-dendrite calcium spikes¹⁰, to more spatially heterogeneous processes such as partial bAP propagation^{13,14} and local spike generation^{15–17}. Such events can provide amplification of synaptic input^{15–17} and the depolarization necessary for Hebbian plasticity induction^{13,18,19}, both of which may be important for place field firing²⁰. However, no measurements of regenerative dendritic activity in place cells have been made during behaviour, when network states affecting dendritic excitability are intact and relevant.

To study regenerative dendritic activity in the hippocampus during behaviour we co-acquired time-series movies through a chronic imaging window of calcium transients from dendrites, axons and somata of CA1 place cells sparsely labelled with a genetically-encoded calcium indicator²¹ (GCaMP6f) while head-restrained mice navigated a virtual linear track^{1,22} (Fig. 1a, b). One imaging plane was focused on the soma while the other was focused in the dendritic arbour, slicing through several branches. Many of the labelled neurons were identified as place cells by somatic calcium transients repeatedly occurring during traversals of the same track location (place field significance $P < 0.05$ from bootstrapping). Unless otherwise stated, our analysis focused only on these cells (33 place fields, 28 place cells, 19.3 ± 13.2 min per place cell imaging session, 8 mice), their basal arbours (170 total branches), their axon (visible in 4 place cells, 5 place fields), and on activity observed during place field traversals. The dendritic fields of view ($\sim 145 \times 75 \mu\text{m}$) on average contained 5 ± 3.5 (range, 2–18) basal arbour branches connected to the

co-imaged place cell soma. The imaged branch sections had a mean length of $10 \pm 3 \mu\text{m}$ (range, 3–23 μm), were positioned a mean of $74 \pm 15\%$ (range, 38–99%) of the distance along the dendritic length (from soma to dendrite tips), a mean of 2.8 ± 0.9 branch points (range, 1–6) and $130 \pm 44 \mu\text{m}$ (range, 58–284 μm) from the soma, and a mean of 3.8 ± 2.0 branch points (range, 1–9) and $191 \pm 83 \mu\text{m}$ (range, 20–440 μm) from each other. Our recordings of ~ 5 basal branches would typically represent

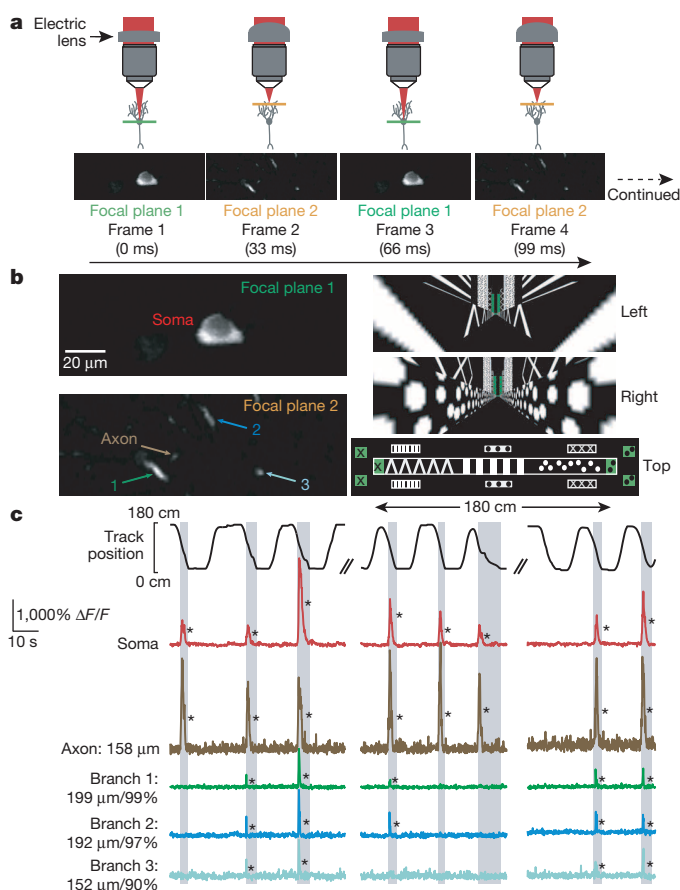


Figure 1 | Co-acquired time-series of CA1 place cell somata, dendrites and axons during virtual navigation. **a**, Two-photon microscopy with an electric lens (top) rapidly switches between two focal planes to generate co-acquired images of soma, dendrites and axon (bottom). **b**, Left, expanded view of place cell shown in **a**. The time-series was acquired while the mouse navigated a virtual linear track (right). **c**, Top, mouse position along linear track; bottom, fluorescence change over baseline ($\Delta F/F$) traces from the soma (red), axon (brown) and dendrites (green, blue, cyan) of place cell shown in **a** during place field traversals (grey columns). Note the absence of detectable branch spikes during some somatic firing. Branch distance to soma and per cent distance from soma to dendrite tip shown for each branch. * $P < 0.001$ from bootstrapping.

¹Department of Neurobiology, Northwestern University, Evanston, Illinois 60208, USA.

about one-third of all basal dendritic branches at their branching depth (~ 2.8 branch points from the soma).

During track traversals we often found significant calcium transients (detectable transients with $<0.1\%$ false positive error rates, indicated with an asterisk in figures; Methods) in the soma, axon (mean of $139 \pm 26 \mu\text{m}$ from soma) and basal dendrites in the cell's somatic place field (Fig. 1b, c; activity in these structures along the track but outside of the place field was rarely observed). Somatic calcium transients (Extended Data Fig. 1) were used as a surrogate measure of action potential firing^{1,21}, and multiple lines of evidence (Methods) showed that somatic action potential firing occurred throughout nearly the entire somatic calcium-transient-defined place field. In cells in which we were able to record from the axon along with the soma and dendrites, we found that axonal transients co-occurred with somatic transients 100% of the time (130 transients; $n = 4$ place cells; Supplementary Video 1). Because axonal calcium transients are tightly coupled to somatic firing, they provided an independent indicator of action potential firing (Methods). In dendrites of CA1 pyramidal neurons of navigating mice (not necessarily place cells), we observed calcium transients restricted to single spines, with no detectable shaft transient; however, we more frequently observed calcium transients invading all visible pixels of the recorded branch (both shaft and spines) that were larger in amplitude and spatial extent than single-spine transients. These transients were considered non-regenerative (excitatory input to a single spine) or regenerative (bAPs or dendritically generated spikes (dspikes) such as Na^+ , NMDA and calcium spikes) depolarizations, respectively (Methods and Extended Data Fig. 2d, e), and we focused our analysis on the latter.

Here, we refer to these regenerative dendritic events as 'branch spikes'—events caused by either dspikes or bAPs or both. Furthermore, because single-spine transients are detectable under our recording conditions, we presume that our measurements are sensitive enough to detect most branch spiking, and therefore the absence of branch spiking additionally indicates that dspikes and bAPs probably did not occur.

The ability to measure spiking in both the soma and in multiple dendrites of CA1 neurons during navigation allowed us to investigate specifically whether somatic place field firing was associated with branch spiking. During many place field traversals with somatic firing, dendritic branch spikes were observed (often with onset latencies with respect to somatic firing, Fig. 2 and Extended Data Fig. 1d), providing the first direct evidence of their existence in place cells. Furthermore, from traversal-to-traversal, branch spiking was found to be highly spatially variable across the arbour. For example, in many place field traversals, all observed dendritic branches in our imaging field (representing a subset of all the cell's branches) spiked along with the soma (Fig. 2a and Extended Data Figs 5 and 6). However, in some cases, only a subset of the observed branches displayed detectable spikes (Figs 1c, 2b and Extended Data Figs 3 and 4a). Finally, in many cases, none of the observed branches displayed detectable spikes during place field traversals while the soma (and axon) fired (Fig. 2c and Extended Data Figs 3–6). From 747 somatic place field transients (from all 28 place cells), 395 (52.9%) showed co-occurring branch spikes in all imaged branches while 154 (20.6%) showed at least one branch with and one branch without detectable spiking, and 198 somatic transients (26.5%) failed to show co-occurring detectable spikes in any of the imaged branches. Importantly, these three observation classes were often seen in the same place cell, but on different place field traversals (Fig. 2e, Extended Data Figs 3 and 4 and Supplementary Video 1).

Characterizing somato-dendritic events as one of these three observation classes did not depend on the recording distance of the dendritic imaging plane from the soma or the number of observed dendritic branches (Supplementary Information and Extended Data Fig. 5). Thus, recording from a subset of branches in a single dendritic imaging plane provides a reasonably accurate characterization of dendritic branch spiking throughout a large portion of the basal (and apical) arbour (Supplementary Information and Extended Data Figs 5 and 6), implying some level of cooperativity between the spiking in different branches (Extended Data Fig. 7).

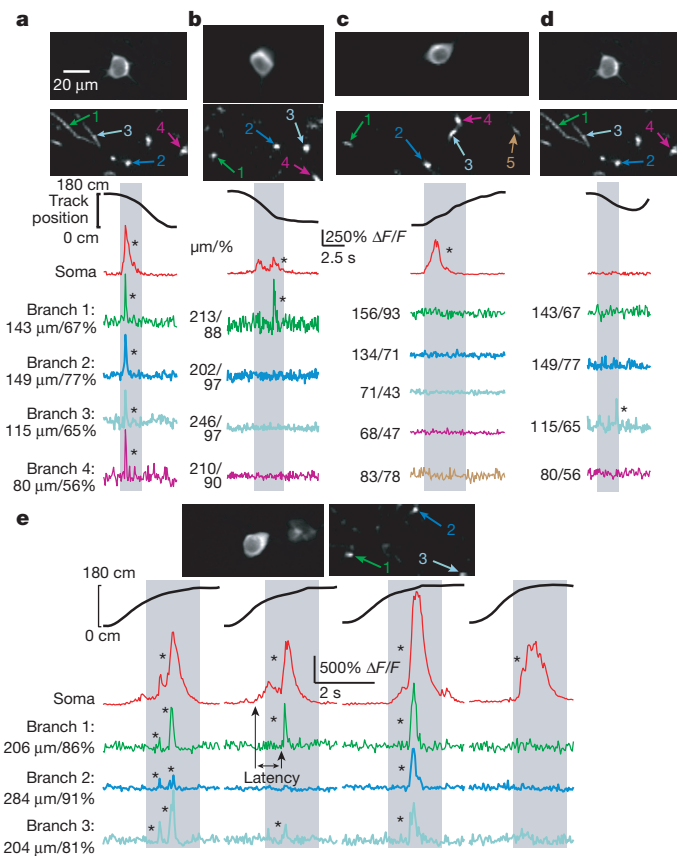


Figure 2 | Variability of dendritic branch spiking during somatic place field firing. **a–d**, Top, co-acquired images of place cell soma and dendrites (**a–c** are different place cells from three mice, **a** and **d** are the same cell). Bottom, $\Delta F/F$ traces from the (co-acquired) soma and numbered dendritic branches during a place field traversal (grey columns). Co-occurring somatic firing and detectable branch spiking was observed in all (**a**), some (**b**) or none (**c**) of the imaged dendrites. Branch spiking in the absence of detectable somatic firing was also observed (**d**). **e**, Variability in branch spiking was often observed in the same place cell (top) during different place field traversals of the same session. Branch spikes also showed variable onset times with respect to somatic firing onset (black arrows). $*P < 0.001$ from bootstrapping.

A fourth class of observation was also made: branch spiking in the absence of detectable somatic spiking (Fig. 2d and Extended Data Fig. 4b), representing probable evidence of dspikes in place cells. These localized dspikes were rarely observed in the absence of somatic spiking (6 total in or around the mean somatic place field, during 8.3 recording hours; 2 other dspikes were observed outside of place fields or in non-place cells; Extended Data Fig. 2a–c, f). However, dspikes could contribute to more global branch spiking, making isolated dspikes difficult to detect.

Place cells exhibited different degrees and patterns of branch-spiking heterogeneity. As a measure of this heterogeneity we calculated the prevalence of branch spiking (fraction of branches with detectable branch spikes) during each place field traversal, defined as branch-spike prevalence (BSP, see Methods for calculation). Figure 3 shows a place field with low average BSP (the average BSP across all place field traversals, Fig. 3a), one with moderate average BSP (Fig. 3b), and one with consistent somato-dendritic unison (high average BSP, Fig. 3c). Further, we observed that average BSP could differ significantly between different place fields of the same cell and that in-field average BSP was significantly greater than out-of-field BSP (Supplementary Information, Extended Data Fig. 8). Together, these results demonstrate that average BSP varies between different place cells and place fields and is not solely determined by cellular properties, such as the general degree of excitability, but is also synaptic-input specific.

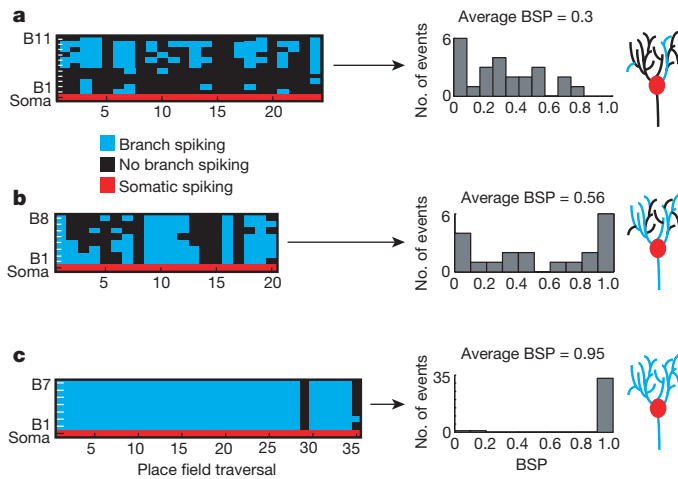


Figure 3 | Place field branch-spiking heterogeneity. **a–c**, Coloured plots (left; three different place cells from two mice) show occurrence of detectable spiking in each branch (blue or black; different branches indicated by 'B' followed by the branch number) during somatic place field firing (red). Middle, histograms of BSP on each traversal. Cartoons (far right) do not represent real data.

The above observations demonstrate spatiotemporal variability in the regenerative dendritic events that are widely believed to provide amplification of synaptic input and the depolarization necessary for Hebbian plasticity (defining plasticity 'windows'). Many possible mechanisms could generate our observed patterns of somatic action potential firing and branch spiking: (1) bAP-induced branch spiking could be spatially and temporally modulated²³ by inhibition, synaptic boosting or sodium channel inactivation; (2) clustered synaptic input could generate branch spiking^{24,25} in the form of widespread calcium spikes and/or more local dspikes^{10,25,26}; or (3) dspikes and bAPs could co-occur. However, regardless of spike initiation site(s), the known involvement of branch spiking in plasticity and input amplification suggests that differences between place cells and their firing fields may be related to BSP.

To investigate the relationship between branch spiking and place field firing, we looked at two features of place fields previously linked to NMDA-receptor mediated plasticity: precision and stability. Place field spatial precision refers to a place cell's somatic spatial firing consistency over many place field traversals, and has been shown to be related to spatial memory task performance^{6,7}. Long-term place field stability, which refers to the persistence of a place field over days, is thought to represent the storage of spatial memories⁴.

We found a significant correlation between a place field's average BSP and its spatial precision index (defined in Methods; Fig. 4a–c; Spearman's rank correlation coefficient: $P = 0.0038$; $\rho = 0.6442$; significant positive linear slope within 95% confidence bounds; 26 place fields). The nine fields with the largest average BSP were also the most precise, indicating that average BSP can be used to predict precision. Notably, when we compared a place field's somatic firing intensity (integral of somatic transients, Methods) to its spatial precision, little to no correlation was observed (Fig. 4c; Spearman's rank correlation coefficient: $P = 0.01$; $\rho = 0.29$; linear slope not significantly positive). Thus average BSP, but not somatic firing intensity, is a predictor of a place field's spatial precision.

We next assessed the relationship between average BSP and place field stability (Fig. 4d–f) by monitoring somatic and dendritic activity in the same cells over the course of 2 days. Place field average BSP was measured on day 1 and then the same cell was imaged the next day to determine whether it had the same somatic place field (Fig. 4d). We classified place fields that persisted to day 2 as stable place fields and those that disappeared by day 2 as transitory place fields (Fig. 4e). Stable fields had significantly greater average BSP than transitory fields (Fig. 4f; 0.83 ± 0.06 versus 0.37 ± 0.07 ; t -test, $P = 0.0006$). When the somatic

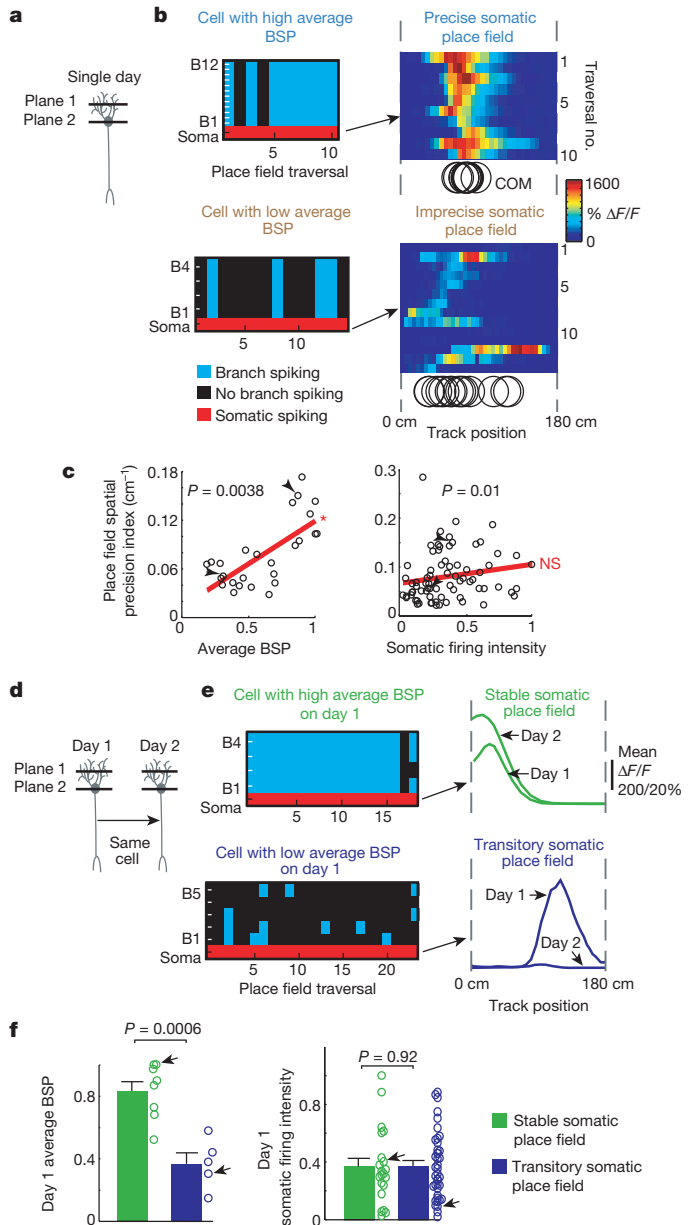


Figure 4 | Dendritic BSP predicts place field spatial precision and long-term stability. **a**, Schematic of imaging planes. **b**, Example cells with high (top left) and low (bottom left) average BSP fields; right, somatic $\Delta F/F$ versus track location raster plots with each traversal's transient centre of mass (COM) (black circles). **c**, Left, plot of precision versus average BSP for each place field ($n = 26$ place fields from $n = 8$ mice); right, plot of precision versus somatic firing intensity (includes single-cell and population imaging experiments; $n = 74$ place fields from $n = 11$ mice). Arrows indicate examples from **b**. NS, not significant. P values from Spearman's rank correlation test; linear fit shown in red; asterisk indicates a significantly positive linear slope. **d**, Schematic of imaging planes on days 1 and 2. **e**, Example cells with high (top left) and low (bottom left) day 1 average BSP; right, associated place fields on days 1 and 2 (scale bar represents 200% and 20% $\Delta F/F$ for top and bottom panels, respectively). **f**, Average BSP (left) and somatic firing intensity (SFI; right, includes single-cell and population imaging experiments) for stable ($n = 8$ place fields from $n = 3$ mice for average BSP group; $n = 21$ place fields from $n = 6$ mice for SFI group) and transitory ($n = 5$ place fields from $n = 2$ mice for average BSP group; $n = 40$ place fields from $n = 5$ mice for SFI group) place fields; error bars represent s.e.m.; P values from Student's unpaired t -test. Black arrows in **f** indicate the example cells shown in **e**.

firing intensity of transitory and stable place fields was compared no significant difference was observed (Fig. 4f; 0.374 ± 0.036 versus 0.368 ± 0.057 ; t -test, $P = 0.92$). This indicates that average BSP, but not somatic firing

intensity, is a predictor of the long-term stability of a place cell's place field. Together, the above results indicate that place field spatial precision, persistence and disappearance can be predicted based on the prevalence of regenerative events in place cell dendrites, but not based on somatic firing alone (precision and stability were also positively correlated, Supplementary Information and Extended Data Fig. 9).

Branch spikes can act to amplify synaptic input (through dspsikes^{15–17}) to drive somatic firing and can also provide the post-synaptic signal required for Hebbian plasticity^{13,18,19}. In this context, our observation that BSP is variable implies that spatially heterogeneous and temporally variable input amplification and synaptic plasticity have a role in place field precision and stability. Somewhat paradoxically, however, increased branch spiking did not strongly correlate with an increased average number of somatic action potentials during place field traversals (Extended Data Fig. 10), and fields associated with more 'open plasticity windows' (more branch spiking possibly associated with greater plasticity) were actually more stable and precise.

It is somewhat puzzling then how synaptic plasticity and amplification mechanisms might be used by place cells to develop their representations of space. A 'winner-takes-all' scenario incorporating synaptic Hebbian potentiation and somatic firing homeostasis provides a possible answer. Previous models^{27,28} examined spatially imprecise place fields generated by different combinations of synaptic input, each with different spatial tuning. The combinations which happened to overlap more (in space and time) generated greater post-synaptic responses and more Hebbian potentiation of the contributing inputs. This increased the effectiveness of the potentiated inputs to drive somatic firing relative to other inputs with non-overlapping spatial tuning. A repetition of this process caused the cell to respond mainly to the subset of strong inputs with spatial tuning overlap, leading to more precise place field firing. The increased effective drive to the cell caused by the potentiated inputs then triggered a homeostatic mechanism that lowered overall cellular excitability (through synaptic or intrinsic excitability renormalization)²⁹. This increased the contribution from the strong inputs (in relation to the weak inputs) in driving somatic firing, thus further increasing spatial precision, but maintaining the average firing rate of the place cell.

Assuming that branch spikes represent the postsynaptic responses that signal Hebbian potentiation, our observation of increased spatial precision of somatic firing with greater average BSP fits with the above model. Also, the sets of surviving, stronger synapses in the model would be expected to persist longer in time, consistent with our observation of greater place field stability with increased average BSP. The use of a somatic action potential firing homeostasis mechanism also explains the weak relationship observed between BSP and action potential firing, with the implication that firing may be driven more by regenerative dendritic activity in high average BSP fields, and less in low average BSP fields. Furthermore, given the above scenario, in silent cells, strong sets of clustered synapses formed through potentiation in different environments may be partially activated in the animal's current environment, but just below dspsike threshold and therefore sensitive to small changes in somatic depolarization, as recently observed³.

In the context of synaptic plasticity (see Supplementary Information for further discussion), while our methods do not directly reveal if branch spikes induce long-term potentiation, long-term depression or no weight change³⁰, the above model implies that in place fields, branch spiking potentiates synapses, is the result of previous potentiation, and/or maintains the strength of previously potentiated synapses through positive feedback. Our measurements demonstrate multiple levels of dissociation between action potential firing and the mechanisms widely believed to signal dendritic plasticity, providing the first clues regarding the spatial and temporal scales at which associative Hebbian learning rules²⁰ may operate in a behaving animal performing a task known to engage integration and plasticity in the cells being studied. One implication is that the process by which place cells orchestrate plasticity signals (dendritic branch spiking) may be variably, rather than consistently, linked to their mode of information transmission (the somato-axonal action potential).

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 29 January; accepted 18 September 2014.

Published online 26 October 2014.

- Dombeck, D. A., Harvey, C. D., Tian, L., Looger, L. L. & Tank, D. W. Functional imaging of hippocampal place cells at cellular resolution during virtual navigation. *Nature Neurosci.* **13**, 1433–1440 (2010).
- Kentros, C. et al. Abolition of long-term stability of new hippocampal place cell maps by NMDA receptor blockade. *Science* **280**, 2121–2126 (1998).
- Lee, D., Lin, B. J. & Lee, A. K. Hippocampal place fields emerge upon single-cell manipulation of excitability during behavior. *Science* **337**, 849–853 (2012).
- Ziv, Y. et al. Long-term dynamics of CA1 hippocampal place codes. *Nature Neurosci.* **16**, 264–266 (2013).
- Mizuseki, K., Royer, S., Diba, K. & Buzsáki, G. Activity dynamics and behavioral correlates of CA3 and CA1 hippocampal pyramidal neurons. *Hippocampus* **22**, 1659–1680 (2012).
- Kentros, C. G., Agnihotri, N. T., Streater, S., Hawkins, R. D. & Kandel, E. R. Increased attention to spatial context increases both place field stability and spatial memory. *Neuron* **42**, 283–295 (2004).
- Rotenberg, A., Mayford, M., Hawkins, R. D., Kandel, E. R. & Muller, R. U. Mice expressing activated CaMKII lack low frequency LTP and do not form stable place cells in the CA1 region of the hippocampus. *Cell* **87**, 1351–1361 (1996).
- Burgess, N. & O'Keefe, J. Neuronal computations underlying the firing of place cells and their role in navigation. *Hippocampus* **6**, 749–762 (1996).
- Solstad, T., Moser, E. I. & Eidevoll, G. T. From grid cells to place cells: a mathematical model. *Hippocampus* **16**, 1026–1031 (2006).
- Grienberger, C., Chen, X. & Konnerth, A. NMDA receptor-dependent multidendritic Ca²⁺ spikes required for hippocampal burst firing *in vivo*. *Neuron* **81**, 1274–1281 (2014).
- Hill, D. N., Varga, Z., Jia, H., Sakmann, B. & Konnerth, A. Multibranch activity in basal and tuft dendrites during firing of layer 5 cortical neurons *in vivo*. *Proc. Natl Acad. Sci. USA* **110**, 13618–13623 (2013).
- Zhou, W. L., Yan, P., Wuskell, J. P., Loew, L. M. & Antic, S. D. Dynamics of action potential backpropagation in basal dendrites of prefrontal cortical pyramidal neurons. *Eur. J. Neurosci.* **27**, 923–936 (2008).
- Magee, J. C. & Johnston, D. A synaptically controlled, associative signal for Hebbian plasticity in hippocampal neurons. *Science* **275**, 209–213 (1997).
- Spruston, N., Schiller, Y., Stuart, G. & Sakmann, B. Activity-dependent action potential invasion and calcium influx into hippocampal CA1 dendrites. *Science* **268**, 297–300 (1995).
- Ariav, G., Polsky, A. & Schiller, J. Submillisecond precision of the input–output transformation function mediated by fast sodium dendritic spikes in basal dendrites of CA1 pyramidal neurons. *J. Neurosci.* **23**, 7750–7758 (2003).
- Gasparini, S., Migliore, M. & Magee, J. C. On the initiation and propagation of dendritic spikes in CA1 pyramidal neurons. *J. Neurosci.* **24**, 11046–11056 (2004).
- Losonczy, A. & Magee, J. C. Integrative properties of radial oblique dendrites in hippocampal CA1 pyramidal neurons. *Neuron* **50**, 291–307 (2006).
- Golding, N. L., Staff, N. P. & Spruston, N. Dendritic spikes as a mechanism for cooperative long-term potentiation. *Nature* **418**, 326–331 (2002).
- Schiller, J., Schiller, Y. & Clapham, D. E. NMDA receptors amplify calcium influx into dendritic spines during associative pre- and postsynaptic activation. *Nature Neurosci.* **1**, 114–118 (1998).
- Wu, X. E. & Mel, B. W. Capacity-enhancing synaptic learning rules in a medial temporal lobe online learning model. *Neuron* **62**, 31–41 (2009).
- Chen, T. W. et al. Ultrasensitive fluorescent proteins for imaging neuronal activity. *Nature* **499**, 295–300 (2013).
- Harvey, C. D., Collman, F., Dombeck, D. A. & Tank, D. W. Intracellular dynamics of hippocampal place cells during virtual navigation. *Nature* **461**, 941–946 (2009).
- Svoboda, K., Denk, W., Kleinfeld, D. & Tank, D. W. *In vivo* dendritic calcium dynamics in neocortical pyramidal neurons. *Nature* **385**, 161–165 (1997).
- Kamondi, A., Acsády, L. & Buzsáki, G. Dendritic spikes are enhanced by cooperative network activity in the intact hippocampus. *J. Neurosci.* **18**, 3919–3928 (1998).
- Smith, S. L., Smith, I. T., Branco, T. & Häusser, M. Dendritic spikes enhance stimulus selectivity in cortical neurons *in vivo*. *Nature* **503**, 115–120 (2013).
- Palmer, L. M. et al. NMDA spikes enhance action potential generation during sensory input. *Nature Neurosci.* **17**, 383–390 (2014).
- Barry, C. & Burgess, N. Learning in a geometric model of place cell firing. *Hippocampus* **17**, 786–800 (2007).
- McHugh, T. J., Blum, K. I., Tsien, J. Z., Tonegawa, S. & Wilson, M. A. Impaired hippocampal representation of space in CA1-specific NMDAR1 knockout mice. *Cell* **87**, 1339–1349 (1996).
- Miller, K. D. Synaptic economics: competition and cooperation in synaptic plasticity. *Neuron* **17**, 371–374 (1996).
- Gordon, U., Polsky, A. & Schiller, J. Plasticity compartments in basal dendrites of neocortical pyramidal neurons. *J. Neurosci.* **26**, 12717–12726 (2006).

Supplementary Information is available in the online version of the paper.

Acknowledgements We are grateful to B. Mensh for discussions on data conceptualization, interpretation and presentation. We thank C. Harvey, B. Mensh, N. Spruston, C. Woolley, W. Kath and L. Looger for comments on the manuscript, E. Han and P. Bueri for technical assistance, and V. Jayaraman, R. Kerr, D. Kim, L. Looger, K. Svoboda from the GENIE Project (Janelia Farm, Howard Hughes Medical Institute) for

GCaMP6. This work was supported by The Klingenstein Foundation, The Whitehall Foundation, The Chicago Biomedical Consortium with support from the Searle Funds at The Chicago Community Trust, Northwestern University, The National Institutes of Health (1R01MH101297), and M.S. is an Ellison Medical Foundation Fellow of the Life Sciences Research Foundation.

Author Contributions M.S. performed the experiments, D.D. built the experimental apparatus, M.S. performed data analysis with strategy suggestions from D.D. Both

authors conceived and designed the experiments, interpreted the data and wrote the paper.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to D.D. (d-dombeck@northwestern.edu).

METHODS

Mouse surgery, virtual reality, and behaviour training. All experiments were approved by the Northwestern University Animal Care and Use Committee. Male C57BL/6 mice (postnatal day ~70) were anaesthetized (~1–2% isoflurane) and a small (~0.5–1.0 mm) craniotomy was made over the hippocampus (1.8 mm lateral, 2.4 mm caudal of Bregma). For single-cell dendritic imaging a low titre Cre virus (AAV2/1–CamkII–Cre, 1.5×10^8 GC ml⁻¹, all virus from University of Pennsylvania Vector Core) was injected (1 injection of ~30 nl at a depth of ~1,250 µm below the dura surface using a bevelled glass micropipette, ~1–2 MΩ after beveling) in combination with a high titre of flexed–GCaMP6 virus (AAV1–Syn–flex–GCaMP6f, 1.4×10^{13} GC ml⁻¹) leading to expression of GCaMP6f (ref. 21) in a sparse CA1 pyramidal neuron population. For population imaging (Extended Data Fig. 9), dense labelling was performed the same as sparse except AAV1–Syn–GCaMP6f (1.5×10^{13} GC ml⁻¹) was injected. Mouse water scheduling began the next day (0.8–1.0 ml per day) followed either the next day or ~7 days later by a hippocampal window and head-plate implantation surgery (as described in ref. 1), for population or single-cell dendritic imaging respectively. Training in a 1.8-m virtual linear track (one ~40–60 min session per mouse per day) began ~7 days after window implantation and continued until mice routinely ran back and forth along the linear track to achieve a high reward rate (>2 rewards per minute); rewards consisted of water (4 µl) delivered as described previously^{1,22}. Once this criterion was reached (~7–14 days of virtual reality training), place cell imaging commenced and was repeated daily for ~1–4 weeks. During the imaging sessions included in our analysis, mice received an average of 4.4 ± 1.1 rewards per min. GCaMP6f expression reached a somewhat steady state level ~21 days post-injection.

Our virtual reality and spherical treadmill system were similar to those previously described^{1,22,31} but with the following differences. A curved screen monitor (CRVD, Ostendo) was used for displaying the virtual reality environment. The screen was adjusted for low light emission (brightness, 15/100; contrast, 50/100; all RGB intensities, 1/100) and was placed ~14 inches in front of the mouse covering 133° and 48° of the mouse's horizontal and vertical field of view, respectively. The Quake2 video game engine described previously²² was used here for the virtual simulation, with minor modifications: the virtual environment was rendered with $2,880 \times 900$ pixels, and the horizontal field of view in the virtual environment was 105°. Mouse locomotion speed and direction on the spherical treadmill were read using a G-400 optical computer mouse (Logitech) and forward and yaw movements of the treadmill were used to update forward position and view angle in the virtual environment, respectively, as described previously²². The forward movement gain was set such that the full length of the virtual track was traversed by ~2.8 rotations of the ball (180 cm of linear distance) and the yaw gain was set such that ~7.5 rotations of the ball resulted in one full field of view rotation (360°) in the virtual environment.

Two-photon imaging of place cell soma, axon and dendrites. We customized a Moveable Objective Microscope (Sutter Instruments) for our imaging experiments. The microscope consisted of water cooled 6215H galvos with 3-mm B1 coated mirrors (Cambridge technology), a $\times 40/0.8$ numerical aperture (NA) objective (LUMPlanFL N $\times 40/0.8$ W, Olympus) and enhanced collection optics. Green GCaMP6f fluorescence was routed to a GaAsP PMT (HI0770PA-40) using a series of dichroic mirrors and band-pass filters (in order after leaving the back aperture; Semrock): FF665-Di02 long pass dichroic, FF01-680/sp. short pass filter, FF560-Di01 long pass dichroic, FF01-510/84 band-pass filter. Stray light from the virtual reality monitor was blocked using a custom box surrounding the top of the microscope objective, the electric lens and the overlying dichroic mirror (not including the tube lens, scan lens, galvanometers or routing mirrors). This box had one hole on top, for entry of the excitation beam, which was covered with a coloured glass filter (FGL780, Thorlabs) and one hole on bottom for the microscope objective. This bottom hole was sealed using the same loose black rubber tube and tight fitting metal rings described previously¹. ScanImage 3.8 was used for microscope control and acquisition³². Ti:Sapphire laser (Chameleon Ultra II, Coherent) light at 920 nm was used as the excitation source. Laser average power at the sample (after the objective) was 10–75 mW. An electric lens³³ (EL-10-30-C-VIS-LD, Optotune; $f = -100$ mm offset lens) was used to switch rapidly between different focal planes (planes up to ~500 µm apart were possible). The electric lens was mounted in close proximity to the microscope back aperture and the focal plane was rapidly switched after each collected image of the time-series by changing the applied steady-state current (LD1255R current driver, Thorlabs). Simultaneous with focal plane switching, the average laser power was modulated, using a pockels cell (350-80-LA-BK-02, 302RM driver, Conoptics), to account for different amounts of tissue penetration (more power for deeper planes).

Images (256×64 pixels, 0.5 ms per line field of view of ~145 \times 75 µm) in each plane were acquired at 31.3 Hz for single-plane, 15.6 Hz for 2-plane and 10.4 Hz for 3-plane acquisitions. The resulting co-acquired time-series movies contain interleaved frames fast enough to provide sufficient sampling of transients in each plane. Note that while the effective frame rate at each plane changed depending on the number of acquired planes, the time to acquire each individual frame (32 ms) did

not change. Time-series imaging sessions lasted 19.3 ± 13.2 min. Potential place cells were initially selected online during behaviour by comparing changes in fluorescence with track location. These cells were then examined to ensure that their dendrites could be clearly separated from dendrites and axons of other cells. Multi-plane imaging was then performed on cells passing these criteria by selecting a dendritic plane(s) with multiple basal branches in the region above (dorsal to) the soma. Time-series acquisition was initiated during behaviour periods of high-reward rate on the linear track task.

A Digidata1440A (Molecular Devices) data acquisition system was used to record (Clampex 10.2) and synchronize position in the linear track, reward timing, and two-photon image frame timing (using the command signal to the slow galvanometer).

After time-series imaging, *z* series were acquired from each cell from the external capsule fibre surface through the proximal apical dendrite ~300 µm deep using the following parameters: 10 frames (short time-series) per *z* plane; 2 µm steps between *z* planes; 256×128 pixels per frame, ~145 \times 75 µm; 1 ms per line. Cell bodies were typically 100–150 µm below the surface.

Data analysis and statistics. Analysis was performed using ImageJ (1.45) and custom scripts written in MatLab (R2011b). All data in the text and figures are presented as mean \pm s.d., except in reference to the plots shown in Fig. 4f, Extended Data Figs 9d and 10b where error represents s.e.m. Sample sizes were chosen to reliably measure experimental parameters while remaining in compliance with ethical guidelines to minimize the number of animals used. Experiments did not involve randomization or blinding because no place cell or animal groups were predefined.

Time-series movies for multi-plane recording were acquired using interleaved frames (that is, every other frame was from the same plane for 2-plane imaging). The electric lens settling time of ~5 ms created distortions in the first few lines of each frame of the movie; these lines were therefore removed before subsequent analysis. Each multi-plane time-series was then split into separate time-series movies, one for each acquired plane. Each single-plane time-series was then independently motion corrected using whole frame cross-correlation, as described previously^{1,34}.

Morphology analysis and identification of axons and dendrites. Because *z* series were acquired during behaviour, motion correction was required to correct for brain movements. Each 10-frame time-series acquired at each *z* plane was motion corrected independently using whole-frame cross-correlation³⁴. A mean image (time projection) was made from each motion-corrected time-series at each *z* plane. The mean images were then registered with respect to each other by cross-correlating³⁴ the mean image at plane 2 with the mean image at plane 1, then plane 3 with 2, and so on until all images were registered.

Dendrite and axon branches were traced (in 19 of 28 place cells; brain motion during high-resolution *z* series prevented accurate reconstruction in 9 cases) from motion-corrected *z* series using Simple Neurite Tracer in Fiji (ImageJ)³⁵. Distances and branch numbers were calculated from the resulting wire segments using custom MatLab scripts. All dendritic and axonal distances from soma represent distance travelled along the neurite; these numbers are provided for each branch in the Figures along with the per cent distance from the soma to the end of the dendrite (dendritic tip).

Dendrites belonging to the co-imaged soma were identified offline by tracing them to the soma in the *z* series; additional verification was provided by their often co-occurring significant calcium transients (see below) with the soma. Axon segments were identified offline by their morphology and signal change. They were smaller in diameter than dendritic branches and also had a lower resting fluorescence level. Furthermore, their calcium transients were longer in duration and larger in signal change compared to dendritic branches.

Region of interest selection and calcium transient analysis. For single-cell imaging (sparse labelling), regions of interest (ROIs) were selected by hand on the mean soma or dendrite images (mean time projection of all frames in the motion-corrected time-series at each plane). ROIs were drawn to closely follow the outline of the structure of interest (soma, axon or dendrite).

For population imaging, ROIs were defined as previously described³⁶ ($\mu = 0.5$, 150 principal components, 100 independent components, s.d. threshold = 1.5, s.d. smoothing width = 1.5, 100 pixels < area of ROI < 400 pixels; see Mukamel *et al.*³⁶ for parameter definitions). As seen previously¹, ROIs nearly always defined single cell regions.

From single-cell and population time-series, $\Delta F/F$ versus time traces were generated for each ROI as previously described¹. In brief, slow changes in the fluorescence traces were removed by examining the distribution of fluorescence in a ~8 s (for single-cell/sparse labelling) or ~3.2 s (for population imaging) interval around each sample in the trace and normalized by the 8th percentile value. These baseline-corrected soma, axon or dendrite fluorescence traces were then subjected to the analysis of the ratio of positive- to negative-deflecting transients of various amplitudes and durations described previously³¹. We used this analysis to identify significant transients with <0.1% false positive error rates; these identified significant transients were used in the subsequent analysis and are marked with an asterisk in

the figures. Note that brain movements can cause small fluorescence transients in the positive- and negative-deflecting directions (equally) during behaviour. Because GCaMP6f indicates activity with positive-deflecting transients only, the negative-deflecting transients are assumed to be due to brain movements. The above cited analysis detects positive-deflecting transients of a duration and amplitude that almost never occurs in the negative-deflecting direction ($<0.1\%$ false positive rate). Thus, most of the small positive-deflecting transients not detected as significant by this analysis are probably induced by brain motion, since events of the same duration and amplitude often occur in the negative-deflecting direction; however some of these small positive-deflecting transients may be due to neural activity (though the number is likely small, see for example our ability to record single-spine events in Extended Data Fig. 2d, e).

The mean duration of somatic, axonal and dendritic calcium transients were calculated as the full duration of the significant transients: soma, 2.89 ± 0.59 s; axon, 2.0 ± 1.2 s; dendrites, 1.07 ± 0.65 s. These durations were calculated from sparsely labelled cells only, where all 3 structures could be identified (not from population labelling). Only transients occurring during consistent runs through the place field, in which a long running period covered $>90\%$ of the place field (see below), were included.

Previous studies have consistently found that somatic calcium transients are caused by action potential firing where the number of underlying action potentials correlates with the change in fluorescence^{21,37–39}. Thus, somatic $\Delta F/F$ fluorescence traces are often used as a surrogate measure of action potential firing activity^{1,22}. The following three related measurements of somatic and axonal calcium transients in the present study are consistent with conclusions of prior studies asserting that somatic calcium transients are caused by action potential firing:

(1) An axonal calcium transient co-occurred with the soma 100% of the time (130 transients; $n = 4$ place cells); axonal transients are tightly coupled to somatic firing^{40,41} and thus provide an independent indicator that action potential firing causes our somatic calcium transients.

(2) Averaging our smallest-amplitude somatic calcium transients produced a trace (Extended Data Fig. 1a) nearly identical in amplitude, shape and duration to what is expected from a single somatic action potential based on previous combined cell-attached and imaging measurements²¹.

(3) Place fields defined by somatic calcium transients are highly similar to those defined by somatic action potential firing in a comparable virtual track^{1,22}.

Taken together, these three points indicate that somatic action potential firing occurs throughout nearly the entire somatic calcium transient-defined place field. Further, somatic calcium transients varied in amplitude, consistent with a difference in the number of underlying action potentials, and varied in duration, consistent with the summation of multiple transients (Extended Data Fig. 1).

Dendritic calcium transients^{42,43} occur as a result of either non-regenerative or regenerative depolarizations that involve calcium influx through local NMDA and voltage-gated calcium channels⁴⁴. Excitatory input to a single spine leads to a non-regenerative post-synaptic depolarization typically detected as a spine-head-restricted (that is, restricted to the head of the spine) calcium transient with little or no shaft component^{21,44–47}. Dendritically generated spikes (such as Na^+ , NMDA and calcium spikes, collectively referred to here as dspikes) and bAPs can also lead to dendritic calcium influx, but by contrast, these events are regenerative depolarizations often detected as both shaft- and spine-invading calcium transients larger in both amplitude and spatial extent than transients associated with single-spine input^{17,44,46,47}. In CA1 pyramidal neurons of navigating mice (not necessarily place cells), we observed calcium transients restricted to single spines, with no detectable shaft transient; however, we more frequently observed calcium transients invading all visible pixels of the recorded branch (both shaft and spines) that were larger in amplitude and spatial extent than single-spine transients (Extended Data Fig. 2d, e). These transients were considered non-regenerative or regenerative depolarizations, respectively. The regenerative dendritic events were defined here as 'branch spikes' — events caused by either dspikes or bAPs or both.

Defining place fields. Place fields were identified and defined as described previously¹, with minor changes outlined below. Place fields were defined solely based on somatic (not dendritic) calcium transients. First, long running periods were defined in each direction in which mouse movement along the virtual track consisted of virtual velocity $>7 \text{ cm s}^{-1}$ and run length $>40 \text{ cm}$ (straight run without changing direction or hitting the end of the track). These long run periods were first categorized based on the running direction (positive or negative direction) and then further subdivided into two categories based on the animal's current task performance. Segments of time between two rewards in which long running periods of only one direction occurred were defined as high-reward-rate periods, all other long running periods were defined as low-reward-rate periods. Only high-reward-rate periods were analysed and all time-series data sets included here had at least 20 (mean of 50) long running segments during high-reward-rate periods in each of the positive and negative directions. For each running direction for each cell, the mean somatic

$\Delta F/F$ was calculated as a function of virtual track position for 80 position bins and this mean fluorescence versus position plot was then averaged over 3 adjacent points. Potential place fields were first identified as contiguous regions of this plot in which all of the points were greater than 25% of the difference between the peak somatic $\Delta F/F$ value (for all 80 bins) and the baseline value (mean of the lowest 20 out of 80 somatic $\Delta F/F$ values). These potential place field regions then had to satisfy the following criteria: 1. The field must be $>13 \text{ cm}$ in width; 2. The field must have one value of at least 10% mean $\Delta F/F$; 3. The mean in field $\Delta F/F$ value must be >3 times the mean out of field $\Delta F/F$ value; and 4. Significant calcium transients must be present $>20\%$ of the time the mouse spent in the place field. Potential place field regions that met these criteria were then defined as place fields if their P value from bootstrapping was <0.05 , as described previously¹ and their mean widths were $<125 \text{ cm}$.

For calculation of BSP, somatic firing intensity, spatial precision and centre of mass (see below) only consistent runs through the place field, in which a long running period covered $>90\%$ of the place field, were considered. Because place field spatial precision might be artificially increased in place fields occurring at the track ends due to edge effects, we only included place fields in which neither edge of the identified field was at a track end. Additionally, we determined that place field spatial precision (for the included fields) does not correlate with distance from track end (not shown).

Place field centre of mass and spatial precision index. To calculate the somatic transient centre of mass (COM) on each traversal (n) along the linear track when a somatic transient occurred in the defined place field, we first split location into 4.5 cm bins (i) and measured somatic $\Delta F/F$ in each bin. We then used the following equation to calculate the COM for each traversal (COM_n):

$$\text{COM}_n = \frac{\sum_i DF_i x_i}{\sum_i DF_i}$$

Where DF_i is the somatic $\Delta F/F$ in bin i and x_i is the distance of bin i from the start of the track. We then calculated the peak $\Delta F/F$ weighted mean COM (COM_w) from all traversals n (COM_n for each traversal was weighted by the peak transient $\Delta F/F$ on that traversal (A_n)):

$$\text{COM}_w = \frac{\sum_n A_n \text{COM}_n}{\sum_n A_n}$$

Spatial precision (SP) was then calculated as the inverse of the peak $\Delta F/F$ weighted COM_w standard deviation as follows:

$$\text{SP} = \frac{1}{\sqrt{\frac{\sum_n A_n (\text{COM}_n - \text{COM}_w)^2}{\sum_n A_n}}}$$

Branch spiking analysis. The images shown (at left) in Figs 3a–c, 4b, e and Extended Data Fig. 8a were generated as follows: somatic and branch spiking events that occurred in the cell's place field (if a cell had more than one place field, each field was treated separately) were converted to ones (a significant transient occurred) and zeros (no significant transient was detected).

Branch-spike prevalence (BSP) was then calculated for each place field traversal with a somatic transient by dividing the sum of the number of branches with a detectable significant transient by the total number of recorded branches. The BSP distribution for each place field in each session could then be examined (Fig. 3a, b, c, Extended Data Figs 7 and 8 histograms) or used to calculate the average BSP for the field (the average of all BSP values for the session). Note that average BSP and resting fluorescence level for each place cell were not statistically significantly related (Extended Data Fig. 5g).

The branch spikes all co-occurred with somatic firing (except where stated otherwise; that is, example in Fig. 2d, and in Extended Data Figs 2 and 4b). All BSP numbers presented are from dendritic activity that occurred during somatic place field firing, except for specified values in Extended Data Fig. 8b (green dots).

To compare in-place field to out-of-place field BSP we examined 5 cells (Extended Data Fig. 8b green dots) that had place fields and also had at least 3 somatic calcium transients occurring randomly along the track (typically during running), but outside of the place field and the reward zones. Note that out-of-place field somatic and dendritic firing along the track (outside of reward zones) was extremely rare.

For each recorded dendritic branch, the onset of each branch spike with respect to the co-occurring somatic transient was measured as the time difference between the onset of the somatic and dendritic significant transients (Fig. 2e and Extended Data Fig. 1d). If multiple branch spikes occurred during the somatic transient, the onset of each branch spike relative to the soma onset was measured.

To calculate how widespread each of the three main types of somato-dendritic observations were (Fig. 2a–c), we classified each event type using one dendritic

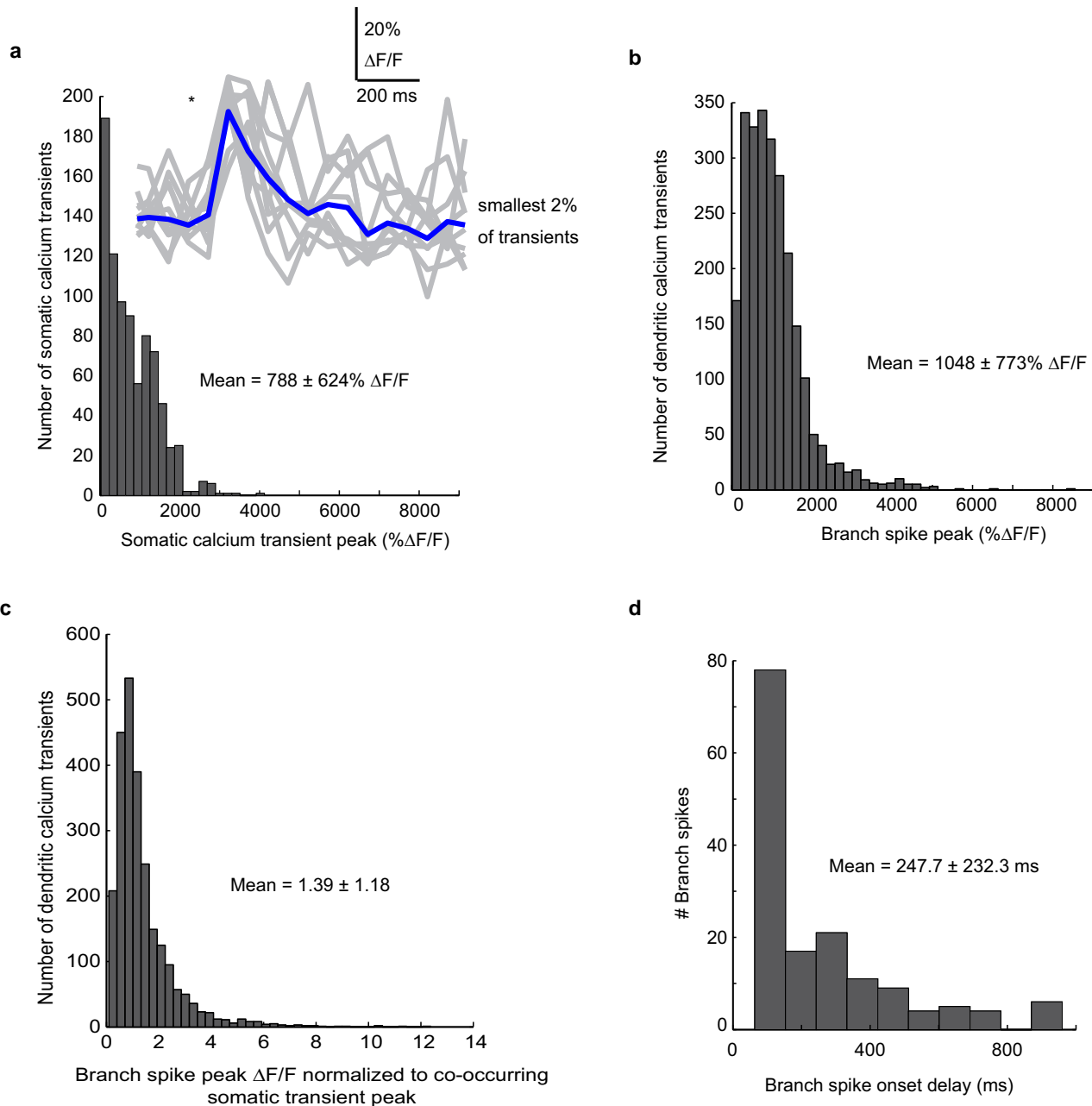
imaging plane and then included all of the branches from a more distal imaging plane (Extended Data Fig. 5a, b) and calculated the percentage of time our classification was correct before and after the distal plane was included.

Somatic firing intensity. The somatic firing intensity is the mean somatic calcium transient integral, calculated by summing the integrals of each somatic calcium transient that occurred in the place field and dividing the sum by the total number of place field somatic calcium transients.

Branch-spike prevalence analysis for stable and transitory place cells. Stable place cells were defined as cells with a mean place field centre of mass (COM_w) on day 1 and day 2 that occurred within 15 cm of each other. In rare cases when place fields shifted between 15 and 50 cm, they were excluded from the analysis. Shifts of >50 cm were considered distinct place fields and so day 1 fields were classified as transitory. Place fields were excluded if they had fewer than 5 place field transients. Day 1 and day 2 were consecutive days. Average branch-spike prevalence and mean somatic firing intensity were measured on day 1 for both stable and transitory place fields.

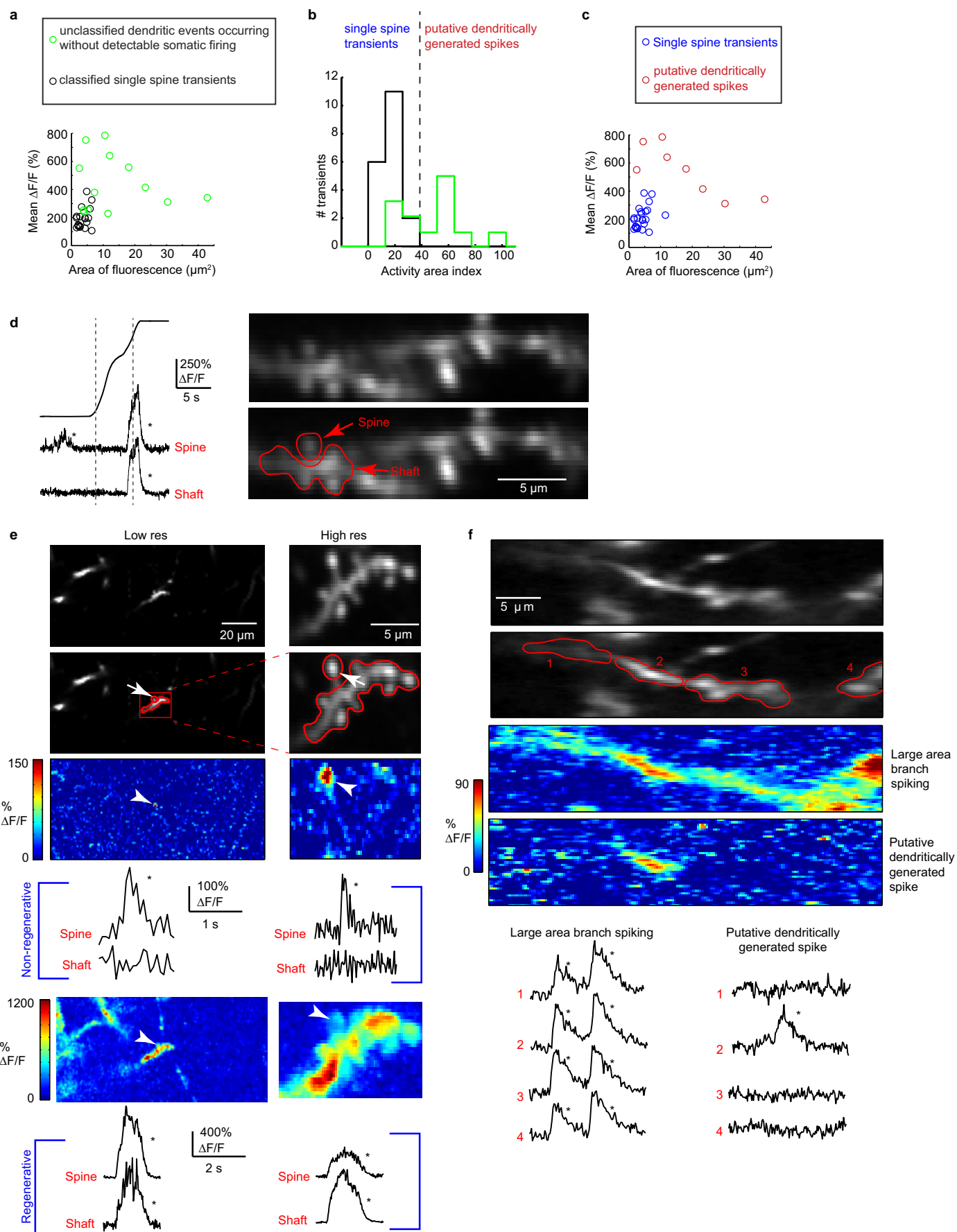
Population analysis for stable and transitory place cells (Extended Data Fig. 9). Place cells were excluded if they had fewer than 5 place field transients, had a place field with one edge at a track end or if the field of view on day 2 did not include the cell. Day 1 and day 2 were consecutive days. Day 8 was seven days after day 1 and no imaging or training took place between day 2 and day 8.

31. Dombeck, D. A., Khabbaz, A. N., Collman, F., Adelman, T. L. & Tank, D. W. Imaging large-scale neural activity with cellular resolution in awake, mobile mice. *Neuron* **56**, 43–57 (2007).
32. Pologruto, T. A., Sabatini, B. L. & Svoboda, K. ScanImage: flexible software for operating laser scanning microscopes. *Biomed. Eng. Online* **2**, 13 (2003).
33. Grewe, B. F., Voigt, F. F., van 't Hoff, M. & Helmchen, F. Fast two-layer two-photon imaging of neuronal cell populations using an electrically tunable lens. *Biomed. Opt. Express* **2**, 2035–2046 (2011).
34. Miri, A., Daie, K., Burdine, R. D., Aksay, E. & Tank, D. W. Regression-based identification of behavior-encoding neurons during large-scale optical imaging of neural activity at cellular resolution. *J. Neurophysiol.* **105**, 964–980 (2011).
35. Longair, M. H., Baker, D. A. & Armstrong, J. D. Simple Neurite Tracer: open source software for reconstruction, visualization and analysis of neuronal processes. *Bioinformatics* **27**, 2453–2454 (2011).
36. Mukamel, E. A., Nimmerjahn, A. & Schnitzer, M. J. Automated analysis of cellular signals from large-scale calcium imaging data. *Neuron* **63**, 747–760 (2009).
37. Greenberg, D. S., Houweling, A. R. & Kerr, J. N. Population imaging of ongoing neuronal activity in the visual cortex of awake rats. *Nature Neurosci.* **11**, 749–751 (2008).
38. Kerr, J. N., Greenberg, D. & Helmchen, F. Imaging input and output of neocortical networks in vivo. *Proc. Natl Acad. Sci. USA* **102**, 14063–14068 (2005).
39. Tian, L. *et al.* Imaging neural activity in worms, flies and mice with improved GCaMP calcium indicators. *Nature Methods* **6**, 875–881 (2009).
40. Cox, C. L., Denk, W., Tank, D. W. & Svoboda, K. Action potentials reliably invade axonal arbors of rat neocortical neurons. *Proc. Natl Acad. Sci. USA* **97**, 9724–9728 (2000).
41. Petreanu, L. *et al.* Activity in motor-sensory projections reveals distributed coding in somatosensation. *Nature* **489**, 299–303 (2012).
42. Helmchen, F., Svoboda, K., Denk, W. & Tank, D. W. In vivo dendritic calcium dynamics in deep-layer cortical pyramidal neurons. *Nature Neurosci.* **2**, 989–996 (1999).
43. Mittmann, W. *et al.* Two-photon calcium imaging of evoked activity from L5 somatosensory neurons in vivo. *Nature Neurosci.* **14**, 1089–1093 (2011).
44. Higley, M. J. & Sabatini, B. L. Calcium signaling in dendrites and spines: practical and functional considerations. *Neuron* **59**, 902–913 (2008).
45. Chen, X., Leischner, U., Rochefort, N. L., Nelken, I. & Konnerth, A. Functional mapping of single spines in cortical neurons in vivo. *Nature* **475**, 501–505 (2011).
46. Sabatini, B. L., Oertner, T. G. & Svoboda, K. The life cycle of Ca^{2+} ions in dendritic spines. *Neuron* **33**, 439–452 (2002).
47. Yuste, R. & Denk, W. Dendritic spines as basic functional units of neuronal integration. *Nature* **375**, 682–684 (1995).
48. Göbel, W., Kampa, B. M. & Helmchen, F. Imaging cellular network dynamics in three dimensions using fast 3D laser scanning. *Nature Methods* **4**, 73–79 (2007).



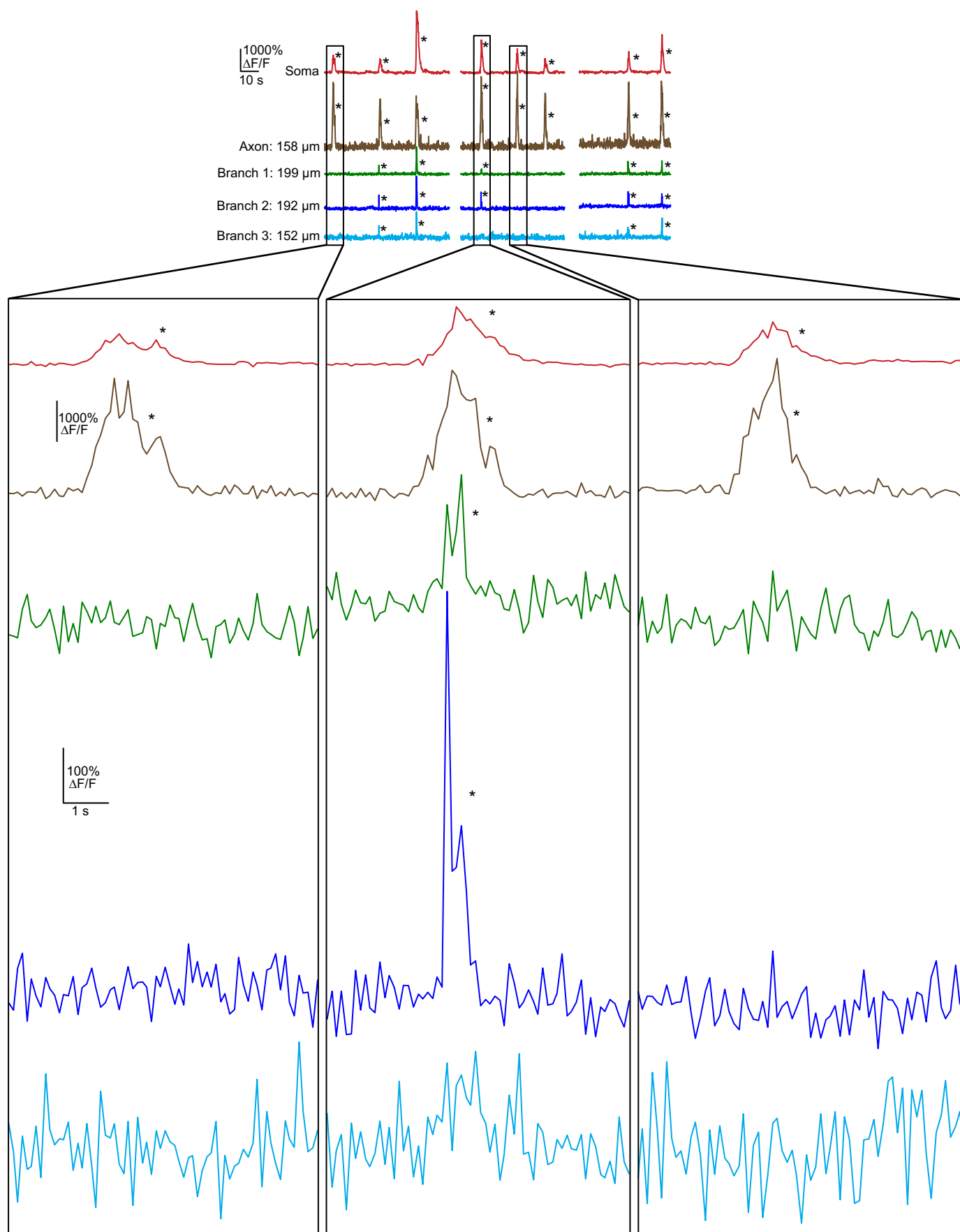
Extended Data Figure 1 | Somatic and dendritic place field Ca^{2+} -transient amplitudes and onset times. **a–c**, Histograms of somatic transient peak $\Delta F/F$ (**a**), branch-spike peak $\Delta F/F$ (**b**) and ratio of branch-spike peak $\Delta F/F$ to somatic transient peak $\Delta F/F$ (**c**). Inset in **a** shows the smallest significant amplitudes of somatic transients detected (grey) and their mean (blue; average triggered when grey transients were first >2 s.d. above the baseline). These small transients are nearly identical in amplitude, shape and duration to what is

expected from single somatic action potentials based on previous *in vivo* combined cell-attached and imaging measurements in the visual cortex using GCaMP6f (ref. 21). **d**, Histogram showing the distribution of branch-spike onset time relative to somatic firing onset. Note that branch-spike onset leading somatic firing onset was not observed, and when branch spiking occurred in multiple branches, their onsets were nearly always simultaneous with respect to each other. Mean \pm s.d. is shown for **a–d**.



Extended Data Figure 2 | Identifying putative dspikes and discriminating between regenerative and non-regenerative dendritic events. **a**, Mean $\Delta F/F$ of significant calcium transients localized to a single branch (using the same imaging parameters to measure co-occurring somatic firing and branch spikes in Figs 1–4), plotted against the area of significant $\Delta F/F$ increase (>3 s.d.). The mean $\Delta F/F$ and the area of significant fluorescence change was calculated for each transient as follows. $\Delta F/F$ movies were generated where each pixel value in each frame of the movie represents the change in fluorescence with respect to the baseline mean for that pixel. The frames during the transient of interest were averaged together and the number of pixels with $\Delta F/F > 3$ were counted, converted to μm^2 and used as the area of significant fluorescence change. The mean $\Delta F/F$ value for the transient was then calculated as the mean value of the pixels with $\Delta F/F > 3$. Black circles represent known single-spine calcium transients acquired using low resolution time-series acquisition (the same resolution used to identify branch spiking), but confirmed as spines using higher-resolution time-series acquisitions where calcium transients were restricted to the spine head (see **d** and **e**). Green circles represent calcium transients restricted to a single dendritic branch and occurring in the absence of somatic firing. Because no high-resolution time-series were acquired from these structures, it was unknown whether they represent transients restricted to single spines or branch spiking. The panels and analysis presented in **b** and **c** indicate that a majority of these events are due to branch spiking and not single-spine transients. **b**, As a combined metric of mean $\Delta F/F$ and area of fluorescence change, we normalized mean $\Delta F/F$ and area of fluorescence to their maximums and measured the distance from the origin (the normalized Euclidean distance of each point in **a**); we refer to this metric as the activity area index (AAI). Histogram showing that known spine transients all fall into the

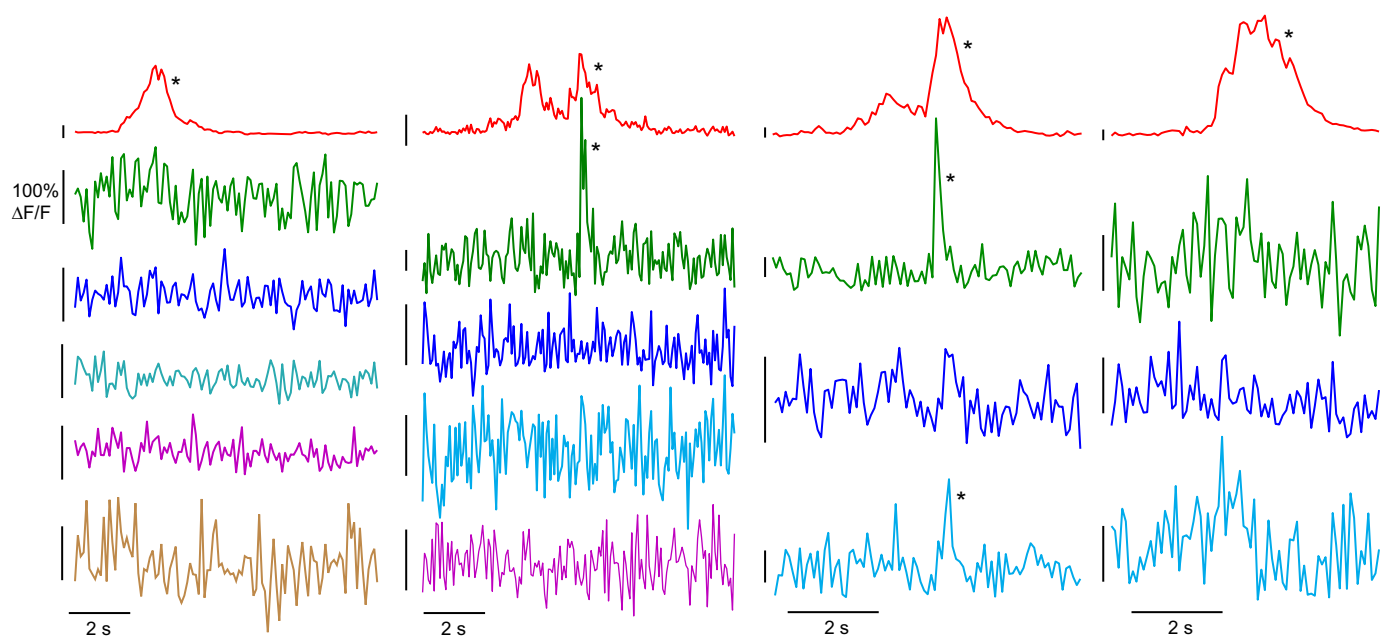
lowest AAI bins (black bars; $\text{AAI} < 40$), and most unclassified events (putative dspikes) in **a** have higher AAIs (green bars). The events in the larger AAI bins (with greater mean $\Delta F/F$ covering a larger area; $\text{AAI} > 40$; separated from the lower AAI bins by the dashed line) fit known characteristics of dspikes. **c**, Using the AAI threshold defined in **b**, most of the unclassified transients fall into a separate group (red) from the known single-spine calcium transients (8 of the 13 unclassified transients had distinctly larger AAIs compared to spine transients, blue) and were therefore considered branch spikes (putative dspikes). **d**, Example calcium transients restricted to a single spine head (bottom left) and invading both spine head and shaft (bottom right) in a place cell. Mean somatic place field is indicated by the grey dashed line. **e**, Example of a stretch of dendrite imaged at low (left) and high (right) resolution in a navigating mouse. The red box and ROIs indicate the same structures that were imaged at both low and high resolution. The same spine head is indicated by arrows at different resolutions in all images. Calcium transients restricted to the same single spine head are shown at both low and high resolution by the colour-coded per cent $\Delta F/F$ map and by the per cent $\Delta F/F$ traces labelled non-regenerative; note that the shaft ROI includes other non-active spines. Calcium transients invading the branch and all spines are shown at both low and high resolution by the colour-coded per cent $\Delta F/F$ map and by the per cent $\Delta F/F$ traces at the bottom, labelled regenerative. **f**, Image of a dendritic branch split into four ROIs. Calcium transients are seen in all parts of the branch during large area branch spiking (see both colour-coded map and traces of per cent $\Delta F/F$). A putative dspike, during navigation, in the same branch causes a significant increase in fluorescence in only part of the branch (ROI2), which includes the shaft.



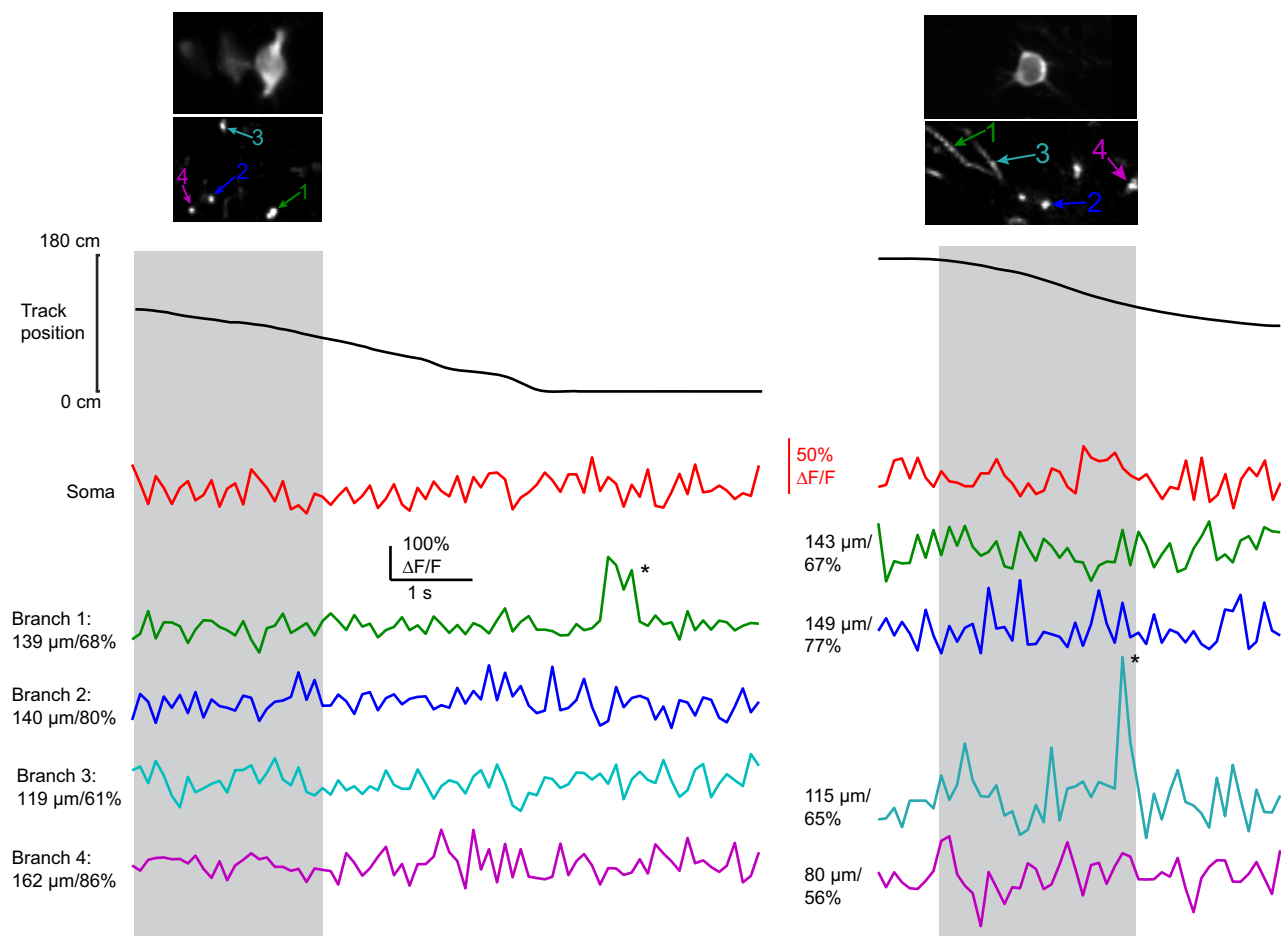
Extended Data Figure 3 | Expanded views of traces showing the variability in detectable events in the soma, dendrites and axon. Traces from Fig. 1b showing variable branch-spike prevalence in the same cell. Three events are

shown amplified (for the amplified traces, the bottom y-axis scale bar refers to the green, blue and cyan dendritic traces, and the top scale bar refers to the red and brown soma and axon traces).

a

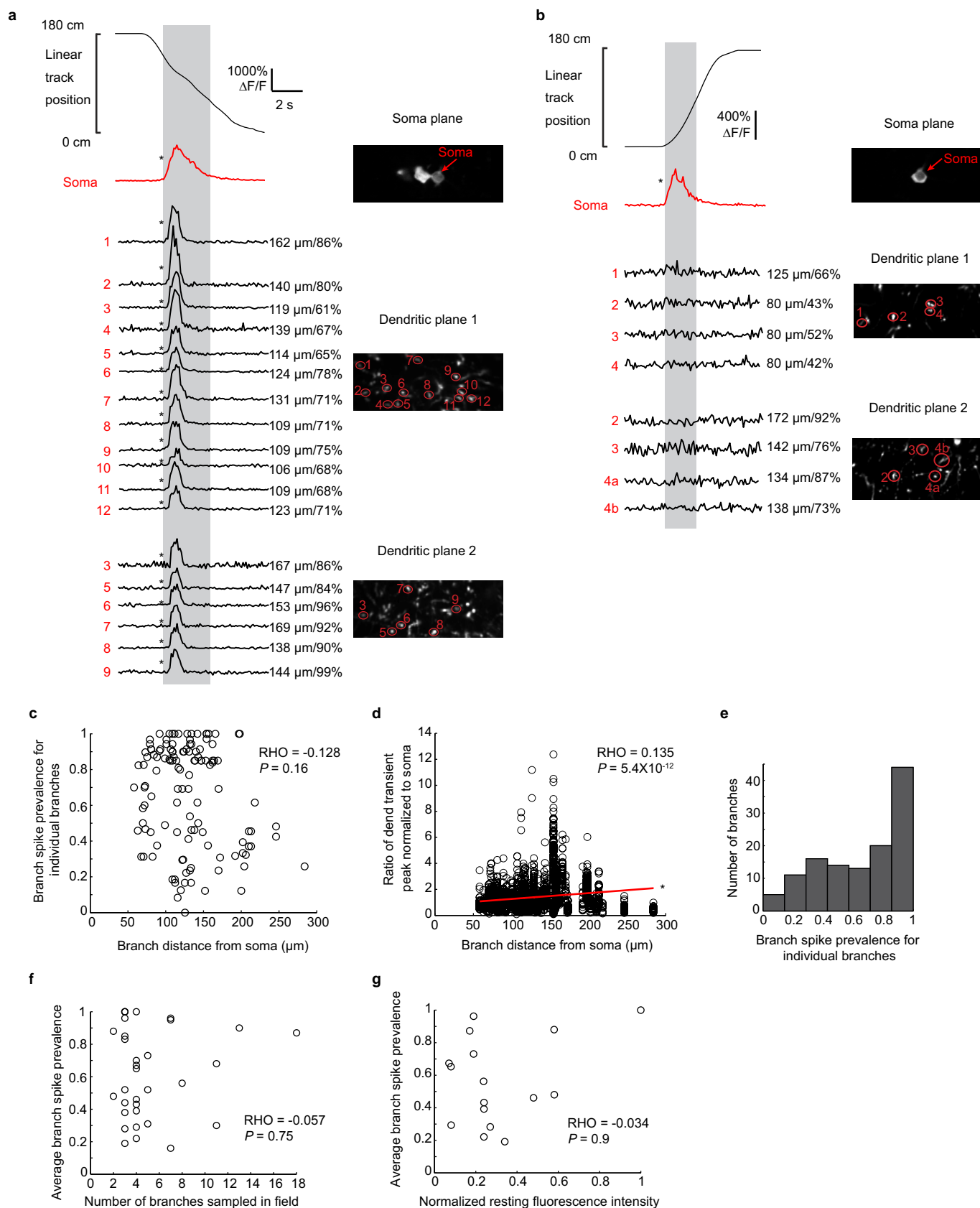


b



Extended Data Figure 4 | Expanded views of traces showing variability of dendritic branch spiking during somatic place field firing. **a**, Traces from Fig. 2 are shown here amplified (each trace has a y-axis scale bar representing 100% $\Delta F/F$). **b**, Two examples of branch spikes detected in single branches in

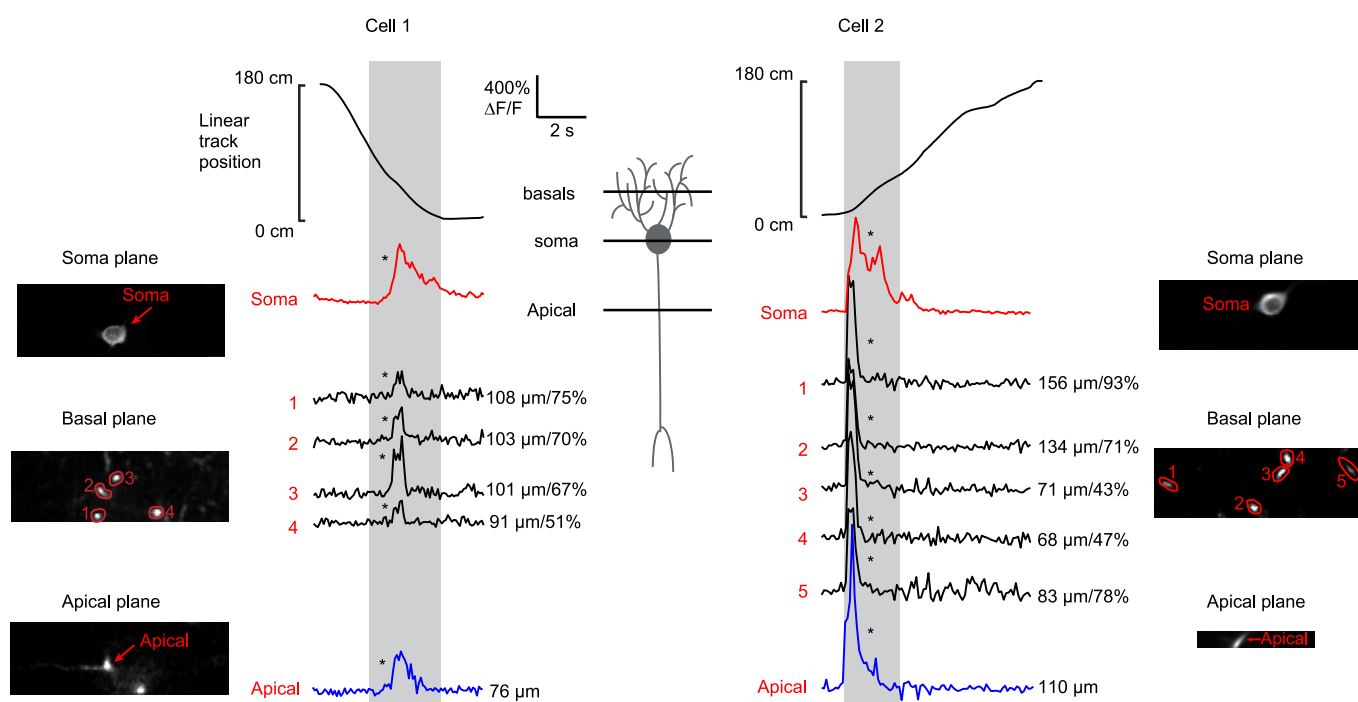
the absence of detectable somatic firing and branch spiking in other dendrites. In each case the soma y-axis is amplified relative to the dendrite traces to show the absence of detectable somatic transients.



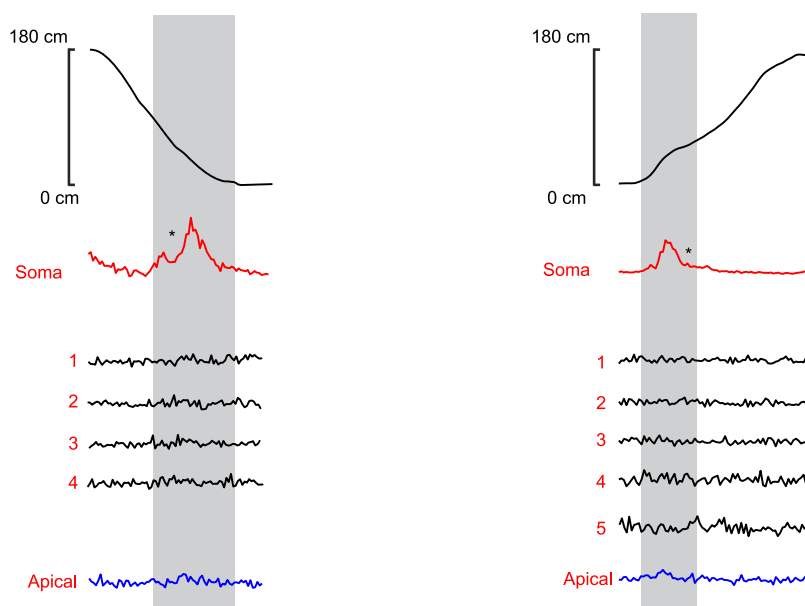
Extended Data Figure 5 | Branch spiking in a single dendritic plane is representative of activity throughout a large portion of the dendritic arbour and average branch-spike prevalence is independent of the number of sampled branches in the field, their distance from the soma and the resting fluorescence level of the soma. **a**, An example of spiking throughout all imaged branches during 3-plane imaging. Co-acquired somatic and dendritic time-series were recorded using three planes; dendritic plane 1 is approximately mid distance between the soma and dendrite tips, dendritic plane 2 is near the branch tips. Numbered arrows indicate branches connected to the imaged soma with the same numbers in dendritic plane 1 and 2 indicating the same branch. An example run through the cell's somatic place field (grey) and corresponding $\Delta F/F$ traces from the soma and numbered branches from the two dendritic planes is shown. Distance along dendrite between branch and soma, and per cent distance from soma to dendritic tip, are on the right of each trace. **b**, Same as **a**, except a different cell where branch spiking is absent throughout all imaged branches in both dendritic planes. **c**, Scatter plot showing the average branch-spike prevalence for individual branches during somatic firing as a function of branch distance from the soma (each point represents one branch). The average branch-spike prevalence for individual

branches is not significantly related to the distance from the soma (Spearman's rank correlation coefficient: $P = 0.16$; $\rho = -0.128$). **d**, Branch-spike peaks normalized to co-occurring somatic peaks during place field traversals from all place cells and branches plotted against branch distance from soma. Spearman's rank correlation coefficient shows a significant correlation ($P = 5.4 \times 10^{-12}$, $\rho = 0.135$) and a linear fit shows a significant positive slope within 95% confidence bounds. **e**, Histogram showing the branch-spike prevalence for individual dendritic branches taken from all cells. **f**, Average branch-spike prevalence (for each place field) plotted against the number of branches sampled in the imaging field shows no significant correlation (Spearman's rank correlation coefficient: $P = 0.75$; $\rho = -0.057$). **g**, Average branch-spike prevalence (for each place field) plotted against normalized resting fluorescence intensity of the soma. Relative resting fluorescence between cells was calculated by dividing the mean measured fluorescence of each soma (not during transients; excluding nucleus) by the squared laser power arriving at the soma (which was estimated based on the soma depth below the surface⁴⁸). Spearman's rank correlation coefficient shows no significant correlation ($P = 0.9$, $\rho = -0.034$).

Somatic spiking with co-occurring branch spikes

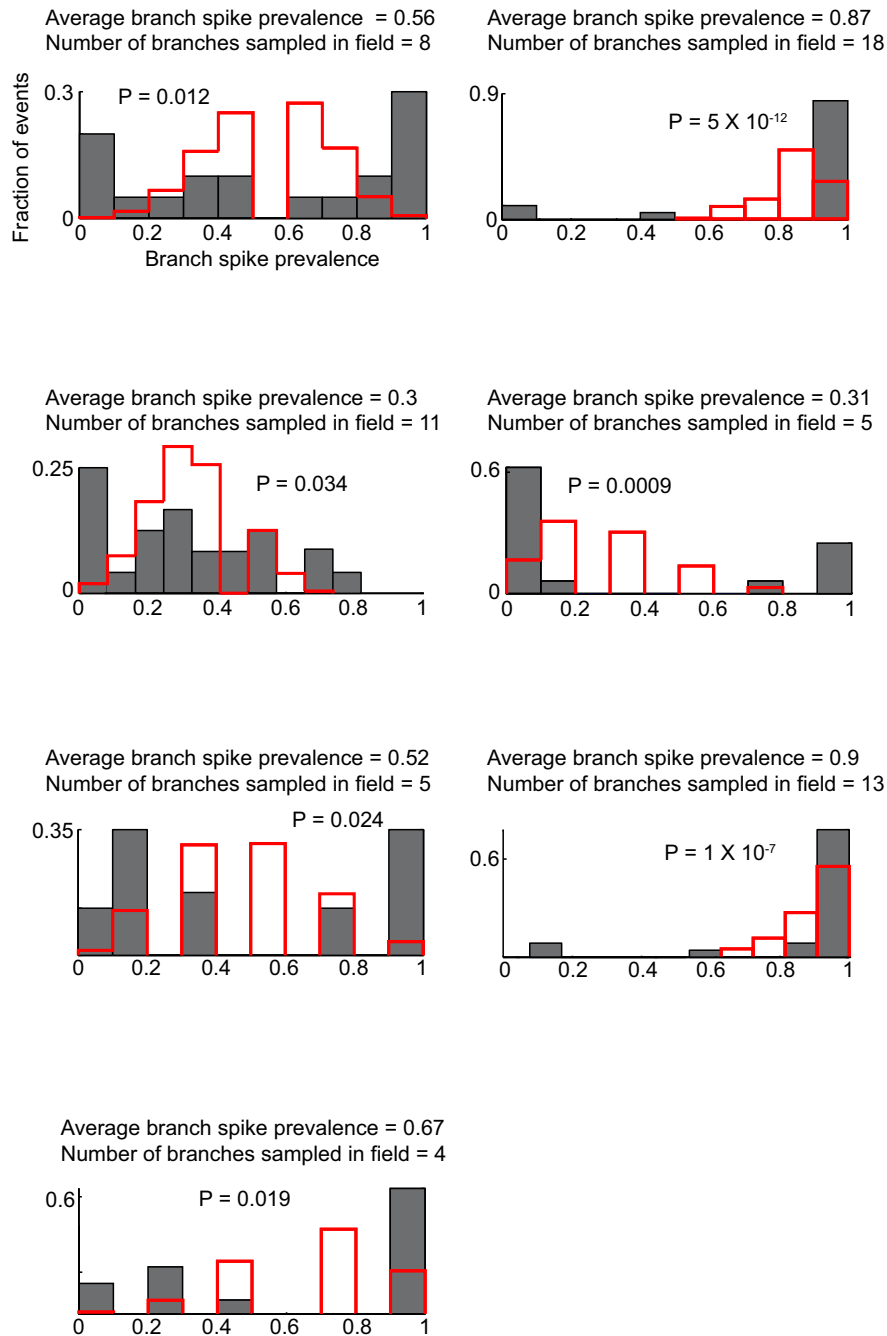


Somatic spiking in the absence of detectable branch spiking in the same cells



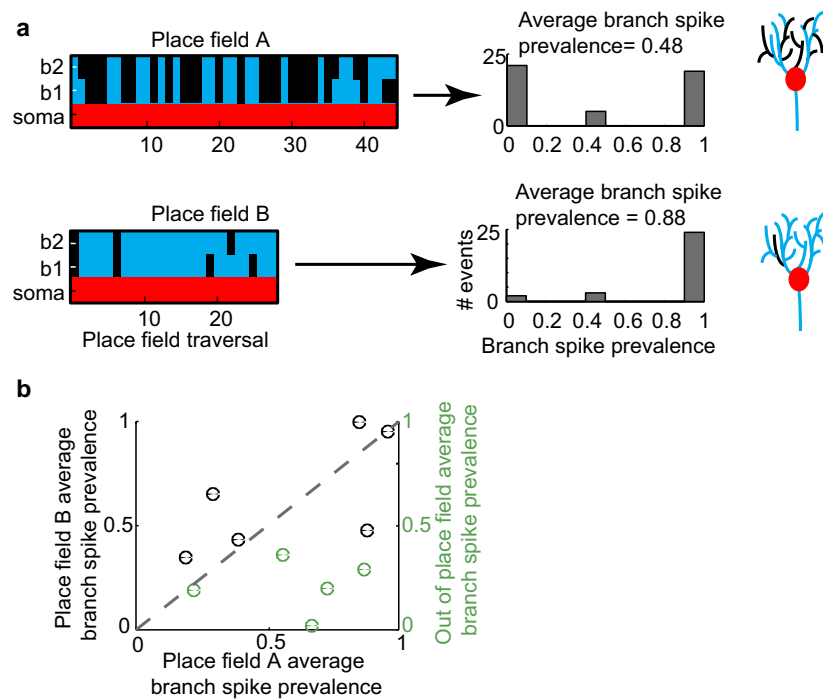
Extended Data Figure 6 | Place field branch spikes in basal and proximal apical dendrites often co-occur. Co-acquired somatic, basal dendritic and apical dendritic (depicted in cartoon in the centre) time-series from two example place cells showing somatic spiking with co-occurring branch spikes (top) and somatic spiking in the absence of detectable branch spikes (bottom)

in the basal and main apical dendrites ($89.6 \pm 19.5\%$ correlation between spiking in the basals and main apical; mean distance of apical site to soma, $109 \pm 38 \mu\text{m}$; $n = 6$ place fields; mean \pm s.d.) during place field traversals (grey). Note that our findings from the basal and proximal apical dendrites may not extend to the oblique dendrites or apical tuft.



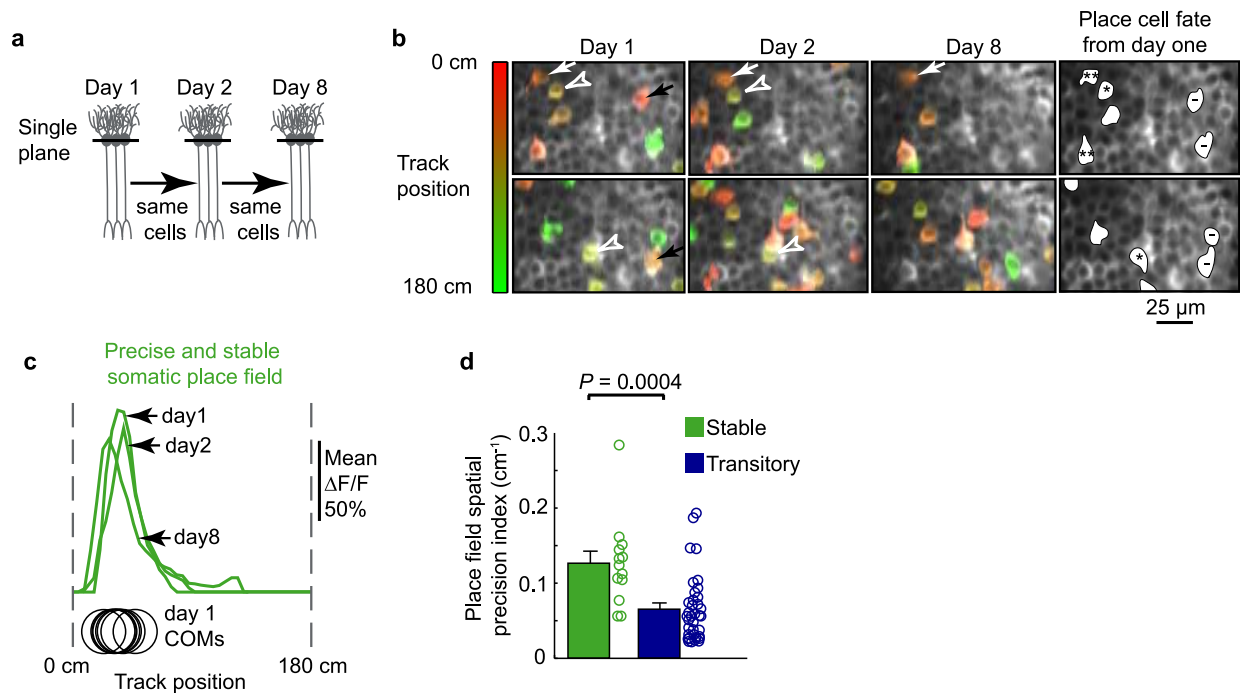
Extended Data Figure 7 | The distribution of branch-spike prevalence differs from a model in which each branch fires independently with a specific probability during place field traversals. Seven histograms from seven example place fields showing the distribution of branch-spike prevalence in each field for real data (grey bars) and modelled data (red bars). The modelled data was generated for each place field example as follows. The probability (P_i) that each dendritic branch (i) in the imaging field would spike in the place field was defined as the branch-spike prevalence for the individual

branch (total number of traversals in which branch i spiked divided by the total number of traversals; from real data). For modelled/mock place field traversals, each branch fired with its random probability P_i and the branch-spike prevalence (fraction of the total number of branches with spikes during the traversal) was calculated. The distribution of branch-spike prevalence was generated for 1,000 modelled/mock place field traversals (red bars). A two-sample Kolmogorov–Smirnov test was used to compare real and modelled distributions (P values shown in each plot).



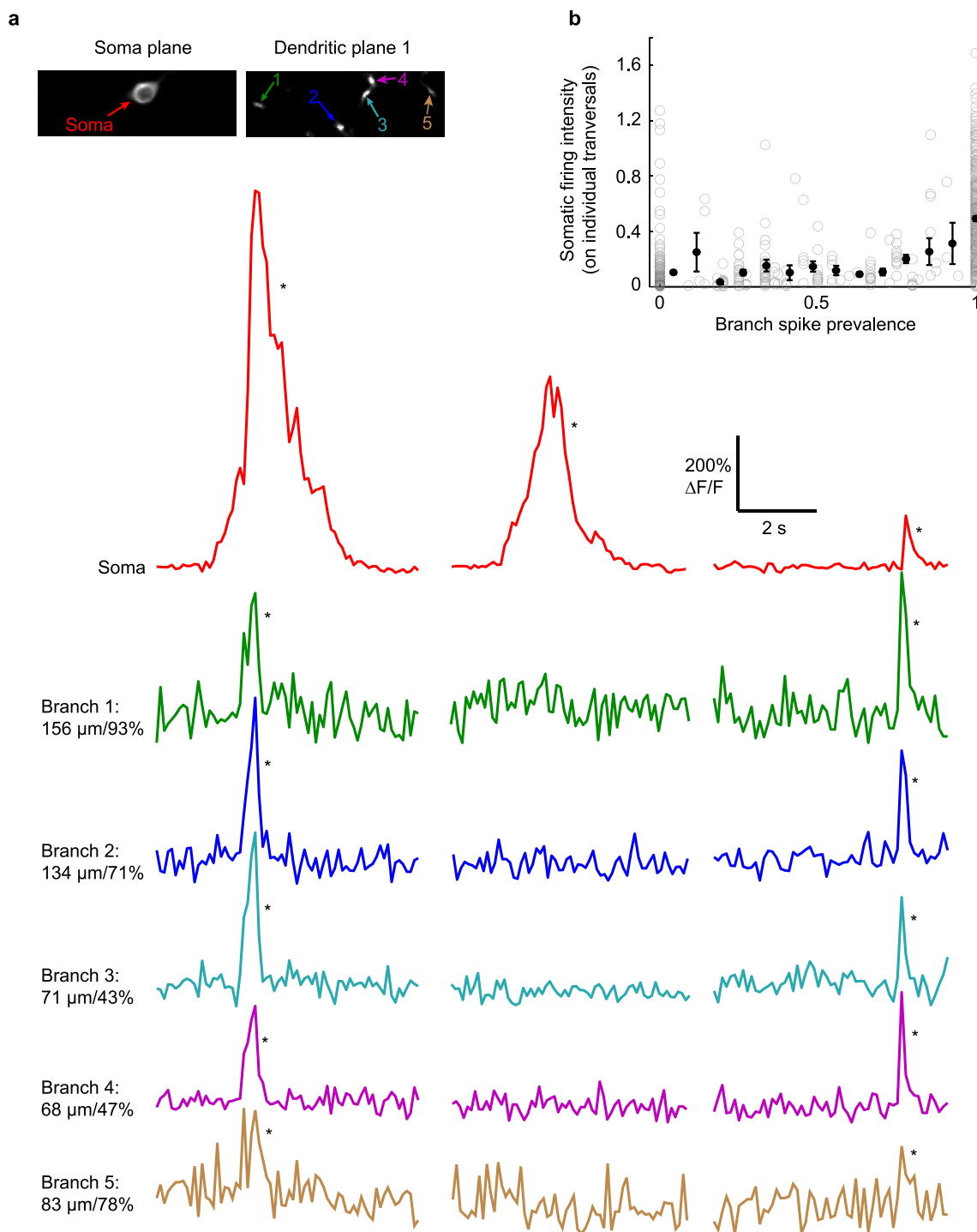
Extended Data Figure 8 | Average branch-spike prevalence can differ between different place fields of the same cell and also between in-place field and out-of-place field somatic firing. **a**, Coloured plots (left) show occurrence of detectable spiking in each branch (blue or black) during somatic place field firing (red) in different co-occurring place fields (A and B, in different running directions) of the same place cell. Right, histograms of branch-spike

prevalence on each traversal for the two place fields. Cartoons (far right) do not represent real data. Note that the running behaviour differed in the two running directions causing differences in the number of traversals reaching behaviour criteria. **b**, Plot comparing average branch-spike prevalence of two distinct place fields in the same place cells (black) or of in-place field versus out-of-place field somatic firing in the same cells (green).



Extended Data Figure 9 | Place field spatial precision is correlated to place field stability. **a**, Schematic showing single-plane imaging of the same population of cell somata over multiple days (1, 2 and 8). **b**, Population images with place cells colour coded by their place field location. Top and bottom rows show place cells in the negative and positive track running directions, respectively. Examples of place cell field fate shown at right: transitory fields occurring on day 1 only (indicated with a minus symbol on the right; black arrows on the left), fields persisting for at least one day (indicated by asterisks

on right, white open arrowheads on left) or seven days (indicated by double asterisks on right; white arrows on left); no symbol indicates that the field did not meet the criteria for inclusion (Methods). **c**, Mean fields of the same cell measured over different days with COM locations from day 1 (bottom; black circles) indicating a precise place field. **d**, Spatial precision is significantly higher in stable versus transitory place fields ($0.127 \pm 0.016 \text{ cm}^{-1}$ versus $0.065 \pm 0.007 \text{ cm}^{-1}$, respectively; t -test; $P = 0.0003$). In **d**, individual data points are depicted by circles to the right of bars and error bars represent s.e.m.



Extended Data Figure 10 | Examples of dissociation between somatic firing and branch spiking in a place cell. **a**, $\Delta F/F$ traces from the (co-acquired) soma and numbered dendritic branches of a place cell during the same imaging session demonstrate that somatic firing and branch spiking are often dissociated. **b**, Plot of somatic firing intensity versus branch-spike prevalence

for all individual place field transversals from all cells (open circles) and binned (solid circles; error bars represent s.e.m.). Branch spiking did not strongly correlate with mean (binned) somatic firing intensity (Spearman's rank correlation coefficient: $P = 0.04$; $\rho = 0.57$).

Precision microbiome reconstitution restores bile acid mediated resistance to *Clostridium difficile*

Charlie G. Buffie^{1,2}, Vanni Bucci^{3,4}, Richard R. Stein³, Peter T. McKenney^{1,2}, Lilan Ling², Asia Gobourne², Daniel No², Hui Liu⁵, Melissa Kinnebrew^{1,2}, Agnes Viale⁶, Eric Littmann², Marcel R. M. van den Brink^{7,8}, Robert R. Jenq⁷, Ying Taur^{1,2}, Chris Sander³, Justin R. Cross⁵, Nora C. Toussaint^{2,3}, Joao B. Xavier^{2,3} & Eric G. Pamer^{1,2,8}

The gastrointestinal tracts of mammals are colonized by hundreds of microbial species that contribute to health, including colonization resistance against intestinal pathogens¹. Many antibiotics destroy intestinal microbial communities and increase susceptibility to intestinal pathogens². Among these, *Clostridium difficile*, a major cause of antibiotic-induced diarrhoea, greatly increases morbidity and mortality in hospitalized patients³. Which intestinal bacteria provide resistance to *C. difficile* infection and their *in vivo* inhibitory mechanisms remain unclear. Here we correlate loss of specific bacterial taxa with development of infection, by treating mice with different antibiotics that result in distinct microbiota changes and lead to varied susceptibility to *C. difficile*. Mathematical modelling augmented by analyses of the microbiota of hospitalized patients identifies resistance-associated bacteria common to mice and humans. Using these platforms, we determine that *Clostridium scindens*, a bile acid 7 α -dehydroxylating intestinal bacterium, is associated with resistance to *C. difficile* infection and, upon administration, enhances resistance to infection in a secondary bile acid dependent fashion. Using a workflow involving mouse models, clinical studies, metagenomic analyses, and mathematical modelling, we identify a probiotic candidate that corrects a clinically relevant microbiome deficiency. These findings have implications for the rational design of targeted antimicrobials as well as microbiome-based diagnostics and therapeutics for individuals at risk of *C. difficile* infection.

Infection with *C. difficile* is a growing public health threat³. Susceptibility to infection is associated with antibiotic use³, and faecal microbiota transplant, which restores microbiota complexity, can resolve recurrent infections⁴. However, the microbiome-encoded genes and biosynthetic gene clusters⁵ critical for infection resistance remain largely undefined, limiting mechanistic understanding and development of microbiota-based therapies. We sought to identify, interrogate, and validate sources of microbiome-mediated *C. difficile* resistance. We first investigated the impact of antibiotics with diverse antimicrobial spectra on the intestinal microbiota and *C. difficile* susceptibility (Extended Data Fig. 1a). Consistent with prior work from our group², administration of clindamycin resulted in long-lasting susceptibility to infection (Fig. 1a). In contrast, ampicillin induced transient susceptibility (Fig. 1c), whereas enrofloxacin did not increase susceptibility to *C. difficile* infection (Fig. 1e). *C. difficile* toxin expression correlated significantly with *C. difficile* abundance in the intestine (Extended Data Fig. 1b). The antibiotic regimens did not substantially alter bacterial density (Extended Data Fig. 1c), but 16S ribosomal RNA (rRNA) gene amplicon sequencing revealed that the three antibiotics had distinct impacts on intestinal microbiota composition (Fig. 1b, d, f).

We exploited this variance in intestinal bacterial composition and infection susceptibility to relate features of microbiota structure to *C. difficile* inhibition. Infection susceptibility correlated with decreased

microbiota alpha diversity (that is, diversity within individuals) (Fig. 2a), consistent with previous studies⁶. Using weighted UniFrac⁷ distances to evaluate beta diversity (that is, diversity between individuals), we found that although clindamycin and ampicillin administration induced distinct changes in microbiota structure, recovery of resistance corresponded with return to a common coordinate space shared by antibiotic-naïve animals (Fig. 2b). However, these diversity metrics generally did not resolve the susceptibility status of animals harbouring microbiota with

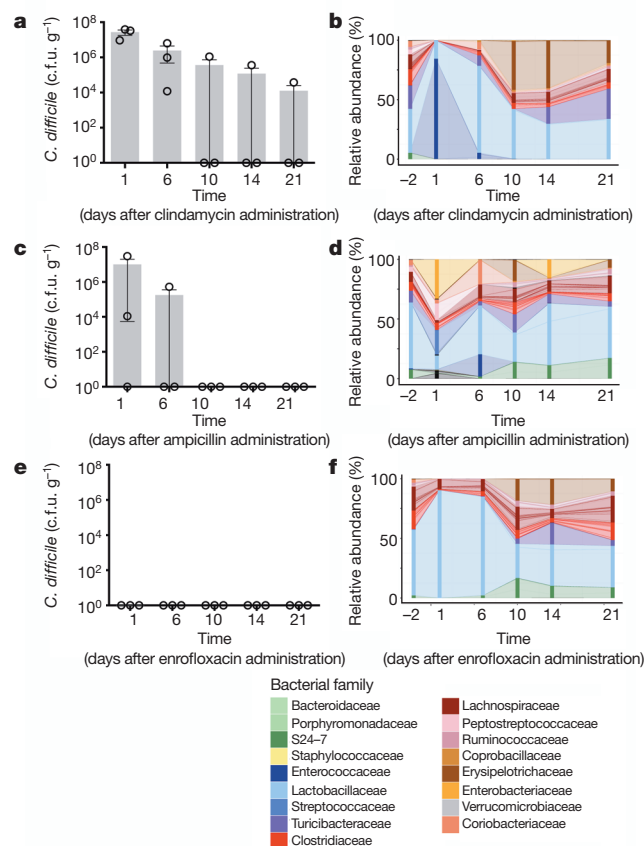


Figure 1 | Different antibiotics induce distinct changes to *C. difficile* infection resistance and intestinal microbiota composition. Susceptibility to *C. difficile* infection after administration of clindamycin (a), ampicillin (c), or enrofloxacin (e). b, d, f, Intestinal microbiota composition at time points indicated. Each stacked bar represents the mean microbiota composition of three separately housed animals. Centre values (mean), error bars (s.e.m.) (a, c, e).

¹Infectious Diseases Service, Department of Medicine, Memorial Sloan Kettering Cancer Center, New York, New York 10065, USA. ²Lucille Castori Center for Microbes, Inflammation and Cancer, Memorial Sloan Kettering Cancer Center, New York, New York 10065, USA. ³Computational Biology Program, Sloan-Kettering Institute, New York, New York 10065, USA. ⁴Department of Biology, University of Massachusetts Dartmouth, North Dartmouth, Massachusetts 02747, USA. ⁵Donald B. and Catherine C. Marron Cancer Metabolism Center, Sloan-Kettering Institute, New York, New York 10065, USA. ⁶Genomics Core Laboratory, Sloan-Kettering Institute, New York, New York 10065, USA. ⁷Bone Marrow Transplant Service, Department of Medicine, Memorial Sloan Kettering Cancer Center, New York, New York 10065, USA. ⁸Immunology Program, Sloan-Kettering Institute, New York, New York 10065, USA.

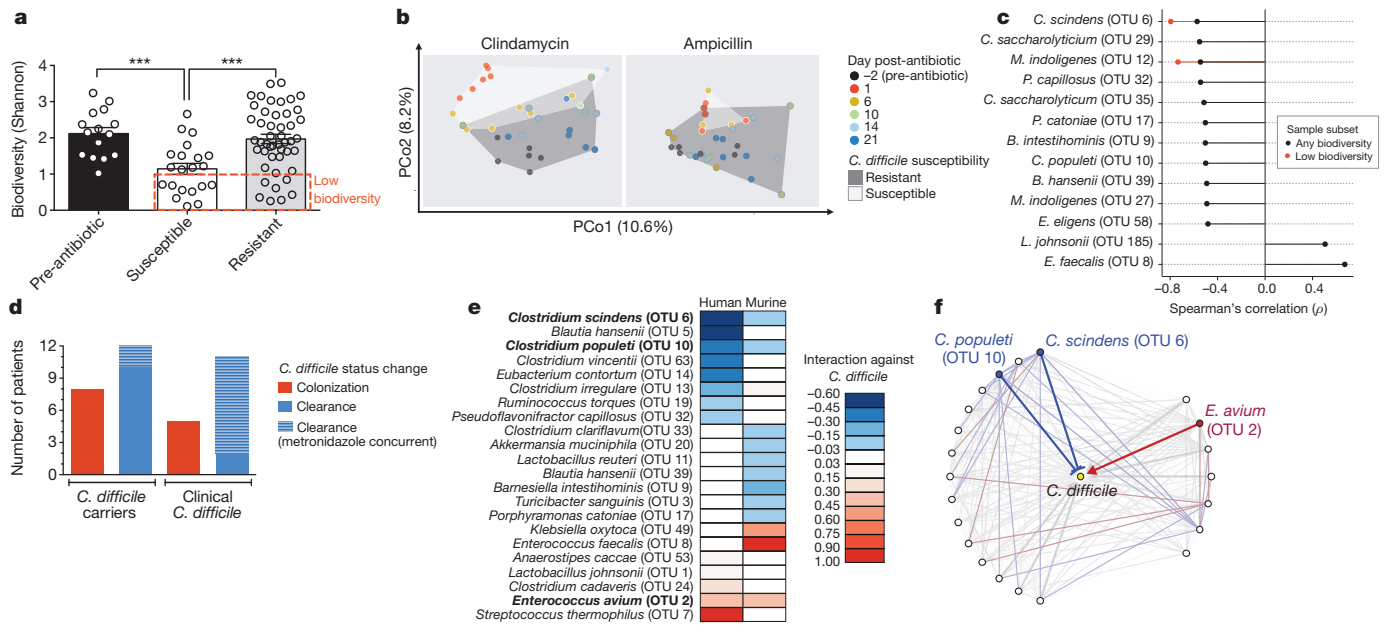


Figure 2 | Native intestinal bacterial species conserved across murine and human microbiota are predicted to inhibit *C. difficile* infection.

Intestinal microbiota alpha diversity (a) and beta diversity (weighted UniFrac distances) (b) of antibiotic-naïve ($n = 15$) and antibiotic-exposed animals susceptible ($n = 21$) or resistant ($n = 47$) to *C. difficile* infection. c, Correlation of individual bacterial OTUs with susceptibility to *C. difficile* infection.

d, Colonization (*C. difficile*-negative to -positive) and clearance (*C. difficile*-positive to -negative) events among *C. difficile*-diagnosed and carrier patients

low alpha diversity (Fig. 2a (red box)) or at early time points after antibiotic exposure (Fig. 2b), suggesting that recovery of more precise microbiota features (for example, individual species) contributed to infection resistance. We correlated resistance with individual bacterial species abundances, corresponding to operational taxonomic units (OTUs, $\geq 97\%$ 16S sequence similarity) (Extended Data Fig. 1d), and identified 11 bacterial OTUs that correlated strongly with infection resistance (Fig. 2c). These OTUs represented a small fraction of the microbiota membership (6%) and comprised primarily *Clostridium* cluster XIVa, including the OTU with the strongest resistance correlation, even among animals harbouring low alpha-diversity microbiota, *C. scindens* (Fig. 2c).

To relate intestinal bacterial species to *C. difficile* resistance in humans, we extended our study to a cohort of patients undergoing allogeneic haematopoietic stem-cell transplantation (allo-HSCT). The majority of these patients were diagnosed with a haematological malignancy and received chemotherapy and/or total body irradiation as well as antibiotics during transplantation (Extended Data Table 1), incurring reduced microbiota biodiversity associated with increased risk of bacterial blood stream infections⁸ and *C. difficile* infection⁹. Compared with controlled animal studies, temporal variation in antibiotic administration and sampling times among patients complicates analysis of relationships between microbiota composition and infection susceptibility. To address these challenges, we employed a recently developed systems biology approach¹⁰ that integrates antibiotic delivery schedules and time-resolved microbiota data to model mathematically the microbiota dynamics and infer which bacteria inhibit *C. difficile*. We included 24 allo-HSCT patients: 12 diagnosed with *C. difficile* infection and 12 who were *C. difficile* carriers without clinical infection (Fig. 2d and Extended Data Fig. 2). To facilitate comparisons across data sets, we clustered murine and human microbiota together to define OTUs that together accounted for a majority of both the human and mouse microbiota structure (Extended Data Fig. 3a–c), and applied the modelling approach to the murine study in parallel. We compared the normalized interaction networks from the human (Extended Data Fig. 3d) and the murine models

included in microbiota time-series inference modelling. Bacterial species with strong *C. difficile* interactions in human and murine microbiota models (e) that exist in a conserved subnetwork predicted to inhibit (blue) or positively associate (red) with *C. difficile* (f). Species interactions in bold type are common to mouse and human. *** $P < 0.001$. In c, $P < 0.0005$ ('any biodiversity', $n = 68$) or $P < 0.05$ ('Low biodiversity', Shannon diversity index ≤ 1 ($n = 16$ animals)). Centre values (mean), error bars (s.e.m.).

(Extended Data Fig. 3e) and identified bacteria displaying strong inhibition against *C. difficile*. Despite some differences across host species networks, the human model identified two *C. difficile*-inhibiting OTUs that were conserved in the murine model (Fig. 2e, f), the strongest of which was *C. scindens*, corroborating our murine correlation-based analyses (Fig. 2c).

To evaluate causality between intestinal bacteria identified in our analyses and infection resistance, we adoptively transferred resistance-associated bacteria. We cultured a representative consortium of four intestinal bacterial isolates with species-level 16S similarity to OTUs associated with *C. difficile* inhibition in our mouse and human analyses (Extended Data Fig. 4) and, after antibiotic administration, animals ($n = 10$) were administered a suspension containing the four-bacteria consortium or vehicle (phosphate-buffered saline (PBS)) before *C. difficile* infection. Additionally, since *C. scindens* had the strongest resistance associations in mice and humans (Fig. 2c, e), we included this bacterium in the consortium and in a third arm alone. Adoptive transfer of the consortium or *C. scindens* alone ameliorated *C. difficile* infection (Fig. 3a, b and Extended Data Fig. 5a) as well as associated weight loss (Fig. 3c and Extended Data Fig. 5b) and mortality (Fig. 3d) significantly compared with control. Transfer of the other three isolates individually, however, did not significantly enhance infection resistance (Extended Data Fig. 5c). Engraftment of the transferred bacteria was confirmed (Extended Data Fig. 5d) by 16S sequence comparison with the input and the native intestinal bacteria from our initial analyses (Fig. 2), thus fulfilling Koch's postulates (albeit for a microorganism and a beneficial health outcome). The abundance of *C. scindens* correlated significantly with infection resistance (Fig. 3e), suggesting that improving bacterial engraftment efficiency may enhance the protective effects of the adoptive transfer. Importantly, bacteria transfer was precise and engraftment did not alter other aspects of microbiota structure compared with control, including density (Extended Data Fig. 5e) and biodiversity (Fig. 3f).

We next interrogated the mechanism of *C. scindens*-mediated *C. difficile* inhibition. Some secondary bile acids can impair *C. difficile*

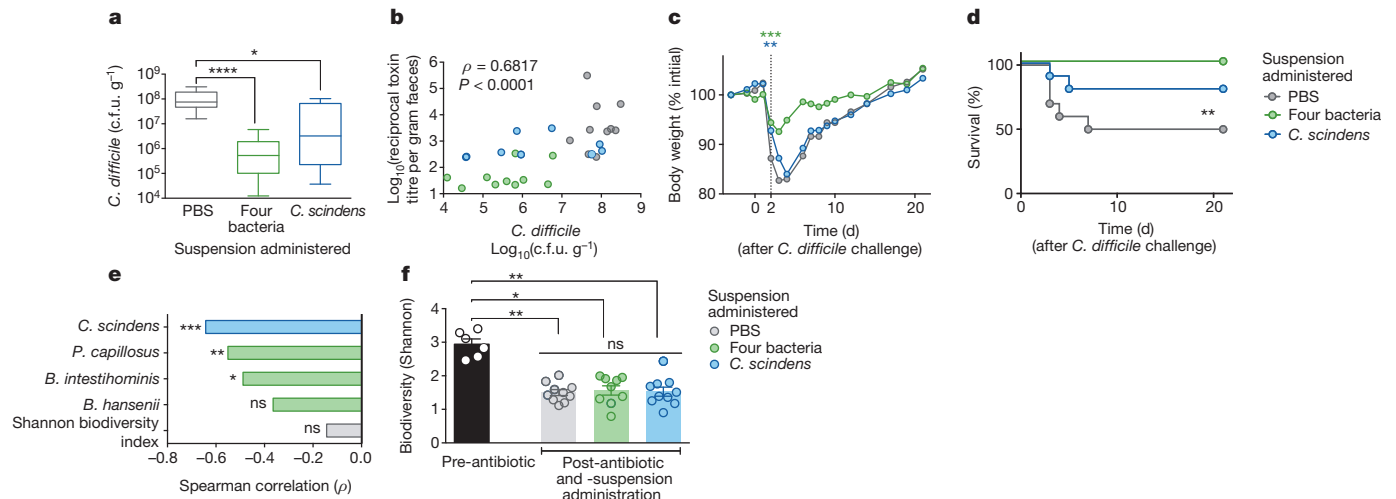


Figure 3 | Adoptive transfer of resistance-associated intestinal bacteria after antibiotic exposure increases resistance to *C. difficile* infection. Intestinal burden of *C. difficile* c.f.u. (a) and toxin (b) 24 h after *C. difficile* infection of antibiotic-exposed animals receiving adoptive transfers. Weight loss (c) and

mortality (d) of animals after infection. e, Correlation of adoptively transferred bacteria engraftment (pre-infection) with *C. difficile* susceptibility. f, Microbiota biodiversity (pre-infection). **** $P < 0.0001$, ** $P < 0.01$, * $P < 0.05$, NS (not significant). Mean (f); error bars, range (a), s.e.m. (f).

growth *in vitro*^{11,12}, but the source and contribution of such metabolites to infection resistance *in vivo* remain unclear. Noting that *C. scindens* expresses enzymes crucial for secondary bile acid synthesis¹³ that are uncommon among intestinal bacteria¹⁴, we hypothesized that the *C. difficile*-protective effects of *C. scindens* may derive from this rare biosynthetic capacity. Analyses of antibiotic-exposed animals (Figs 1 and 2) revealed that recovery of secondary bile acids and the abundance of the gene family responsible for secondary bile acid biosynthesis (as predicted using PICRUSt¹⁵) correlated with *C. difficile* resistance (Fig. 4a, b). Targeted microbiome analysis of the gene family responsible for secondary bile acid biosynthesis indicated that abundance of the bile acid inducible (*bai*) operon genes correlated strongly with resistance to *C. difficile* infection (Fig. 4c) but that bile salt hydrolase (BSH)-encoding gene abundance did not. These results are consistent with reports indicating that BSH-encoding genes are distributed broadly while an extremely small fraction of intestinal bacteria possess a complete secondary bile acid synthesis pathway¹⁴. PCR-based assay of *baiCD*, which encodes

the 7 α -hydroxysteroid dehydrogenase enzyme critical for secondary bile acid biosynthesis, revealed that animals that had recovered *C. difficile* resistance after antibiotic exposure harboured a *baiCD*⁺ microbiome, whereas susceptible animals did not (Extended Data Fig. 6a).

Recipients of either the consortium or *C. scindens* harboured *baiCD*⁺ microbiomes with restored abundance of secondary bile acid biosynthesis genes (predicted by PICRUSt) (Extended Data Fig. 6b). Administration of either bacterial suspension also restored relative abundance of the secondary bile acids deoxycholate (DCA) (Fig. 4d) and lithocholate (LCA) (Extended Data Fig. 7a), both of which inhibit *C. difficile* in a dose-dependent fashion (Extended Data Fig. 8a, b), but abundances of primary bile acids were not significantly altered (Extended Data Fig. 7). Metagenomic inference indicated that consortia recipients harboured microbiomes with greater abundances of secondary bile acid biosynthesis genes than *C. scindens* recipients (Extended Data Fig. 6b), perhaps explaining their superior resistance to *C. difficile*. However, intestinal abundances of DCA and LCA were each comparable in the consortia

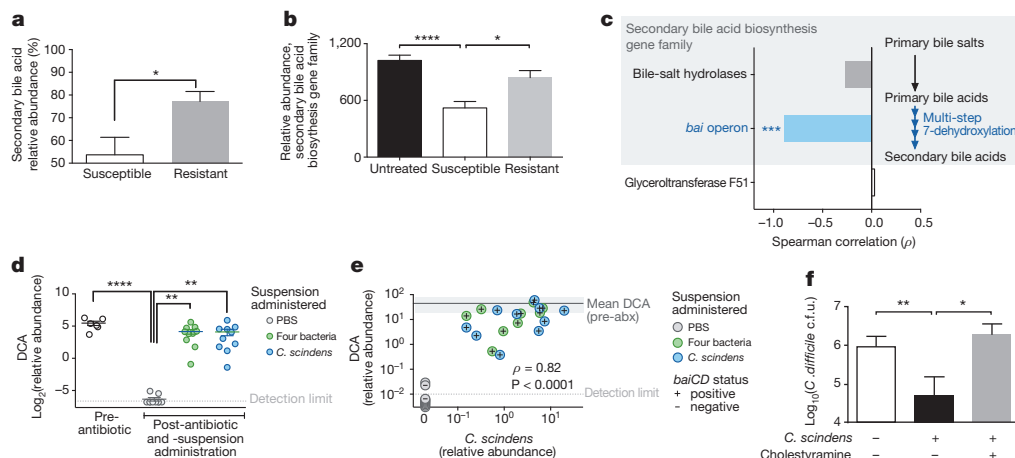


Figure 4 | *C. scindens*-mediated *C. difficile* inhibition is associated with secondary bile acid synthesis and is dependent on bile endogenous to intestinal content. Relative abundance of secondary bile acid species (a) and biosynthesis gene family abundance predicted by PICRUSt (b) in intestinal content from antibiotic-exposed *C. difficile* susceptible ($n = 21$), resistant ($n = 47$), and pre-antibiotic ($n = 15$) animals. c, Correlation of *C. difficile* susceptibility with the abundance of the gene family responsible for secondary bile acid biosynthesis in intestinal content ($n = 6$) quantified using shotgun

sequencing. d, Intestinal abundance of DCA after adoptive transfer of bacteria ($n = 10$ per group). e, Correlation of *C. scindens* engraftment with DCA abundance and *baiCD* status in intestinal content of antibiotic-exposed, adoptively transferred animals ($n = 30$). f, Bile acid dependent *C. scindens*-mediated inhibition of *C. difficile* quantified *ex vivo* ($n = 6$ per group). **** $P < 0.0001$, *** $P < 0.001$, ** $P < 0.01$, * $P < 0.05$. Glyceroltransferase F51, endogenous reference gene (c). Shaded region around 'Mean DCA pre-abx' (pre-antibiotic), DCA abundance (s.d.) (e).

and *C. scindens* recipients pre-infection challenge (Fig. 4d and Extended Data Fig. 7a), suggesting additional mechanisms enhanced colonization resistance in consortia recipients. Indeed, of the four transferred bacteria, only *C. scindens* was *baiCD*⁺ (Extended Data Fig. 6a). Engraftment of *C. scindens* also correlated strongly with DCA relative abundance and *baiCD* in recipients, reaching levels observed in antibiotic-naïve animals (Fig. 4e), which indicated that precise transfer and efficient engraftment of this bacterium could restore physiological levels of secondary bile acid synthesis in antibiotic-exposed animals.

We evaluated bile acid dependent microbiota-mediated inhibition of *C. difficile* using an *ex vivo* model. Pre-incubation of intestinal content from antibiotic-naïve animals with cholestyramine, a bile acid sequestrant¹⁶, permitted *C. difficile* growth (Extended Data Fig. 8c, d) comparable to intestinal content from antibiotic-exposed animals. Consistent with *in vivo* findings, introduction of *C. scindens* significantly inhibited *C. difficile* in the intestinal content from antibiotic-exposed animals. This effect was neutralized when intestinal content was pre-incubated with cholestyramine (Fig. 4f), indicating that *C. scindens*-mediated inhibition of *C. difficile* is dependent upon accessing and modifying endogenous bile salts and recapitulates a natural mechanism of microbiota-mediated infection resistance.

In summary, we show that a fraction of the intestinal microbiota as precise as a single bacterial species confers infection resistance by synthesizing *C. difficile*-inhibiting metabolites from host-derived bile salts. Our use of a human-derived *C. scindens* isolate to augment murine *C. difficile* inhibition emphasizes the conservation of this finding across species and suggests therapeutic and diagnostic applications. The genus *Clostridium* is phylogenetically complex^{17,18}, highlighting the value of integrating functional genomic and metabolomic interrogation with 16S rRNA profiling when evaluating probiotic candidates. Most bile-acid 7-dehydroxylating bacteria are cluster XIVa Clostridia closely related to one another^{14,19,20} and resistance-associated OTUs we identified, suggesting that *bai* or 16S gene signatures may serve as specific, functionally meaningful biomarkers for infection resistance. The replenishment of secondary bile acids and/or biosynthesis-competent bacteria (such as *C. scindens*) may contribute to the therapeutic efficacy of faecal microbiota transplant²¹. Attempts to manipulate intestinal bile acids directly should be performed with caution since some secondary bile acids have been linked to gastrointestinal cancers²². Other bacteria may augment resistance by enhancing 7-dehydroxylating Clostridia or through additional orthogonal mechanisms, such as competition for mucosal carbohydrates²³, activation of host immune defences^{24,25}, or production of antibacterial peptides²⁶. Knowledge of such mechanisms and the ecological context of those microbes responsible will facilitate amplification of microbiota-mediated pathogen resistance in individuals at risk of infection.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 4 May; accepted 3 September 2014.

Published online 22 October 2014; corrected online 7 January 2015 (see full-text HTML version for details).

1. Buffie, C. G. & Pamer, E. G. Microbiota-mediated colonization resistance against intestinal pathogens. *Nature Rev. Immunol.* **13**, 790–801 (2013).
2. Buffie, C. G. *et al.* Profound alterations of intestinal microbiota following a single dose of clindamycin results in sustained susceptibility to *Clostridium difficile*-induced colitis. *Infect. Immun.* **80**, 62–73 (2012).
3. Rupnik, M., Wilcox, M. H. & Gerding, D. N. *Clostridium difficile* infection: new developments in epidemiology and pathogenesis. *Nature Rev. Microbiol.* **7**, 526–536 (2009).
4. van Nood, E. *et al.* Duodenal infusion of donor feces for recurrent *Clostridium difficile*. *N. Engl. J. Med.* **368**, 407–415 (2013).
5. Cimermancic, P. *et al.* Insights into secondary metabolism from a global analysis of prokaryotic biosynthetic gene clusters. *Cell* **158**, 412–421 (2014).

6. Chang, J. Y. *et al.* Decreased diversity of the fecal microbiome in recurrent *Clostridium difficile*-associated diarrhea. *J. Infect. Dis.* **197**, 435–438 (2008).
7. Lozupone, C. & Knight, R. UniFrac: a new phylogenetic method for comparing microbial communities. *Appl. Environ. Microbiol.* **71**, 8228–8235 (2005).
8. Taur, Y. *et al.* Intestinal domination and the risk of bacteremia in patients undergoing allogeneic hematopoietic stem cell transplantation. *Clin. Infect. Dis.* **55**, 905–914 (2012).
9. Kinnebrew, M. A. *et al.* Early *Clostridium difficile* infection during allogeneic hematopoietic stem cell transplantation. *PLoS ONE* **9**, e90158 (2014).
10. Stein, R. R. *et al.* Ecological modeling from time-series inference: insight into dynamics and stability of intestinal microbiota. *PLOS Comput. Biol.* **9**, e1003388 (2013).
11. Wilson, K. H. Efficiency of various bile salt preparations for stimulation of *Clostridium difficile* spore germination. *J. Clin. Microbiol.* **18**, 1017–1019 (1983).
12. Sorg, J. A. & Sonenshein, A. L. Bile salts and glycine as cogerminants for *Clostridium difficile* spores. *J. Bacteriol.* **190**, 2505–2512 (2008).
13. Kang, D. J., Ridlon, J. M., Moore, D. R., Barnes, S. & Hylemon, P. B. *Clostridium scindens* *baiCD* and *baiH* genes encode stereo-specific 7 α /7 β -hydroxy-3-oxo- Δ^4 -cholenic acid oxidoreductases. *Biochim. Biophys. Acta* **1781**, 16–25 (2008).
14. Ridlon, J. M., Kang, D. J. & Hylemon, P. B. Bile salt biotransformations by human intestinal bacteria. *J. Lipid Res.* **47**, 241–259 (2006).
15. Langille, M. G. *et al.* Predictive functional profiling of microbial communities using 16S rRNA marker gene sequences. *Nature Biotechnol.* **31**, 814–821 (2013).
16. Out, C., Groen, A. K. & Brufau, G. Bile acid sequestrants: more than simple resins. *Curr. Opin. Lipidol.* **23**, 43–55 (2012).
17. Collins, M. D. *et al.* The phylogeny of the genus *Clostridium*: proposal of five new genera and eleven new species combinations. *Int. J. Syst. Bacteriol.* **44**, 812–826 (1994).
18. Yutin, N. & Galperin, M. Y. A genomic update on clostridial phylogeny: Gram-negative spore formers and other misplaced clostridia. *Environ. Microbiol.* **15**, 2631–2641 (2013).
19. Kitahara, M., Takamine, F., Imamura, T. & Benno, Y. Assignment of *Eubacterium* sp. VPI 12708 and related strains with high bile acid 7 α -dehydroxylating activity to *Clostridium scindens* and proposal of *Clostridium hylemonae* sp. nov., isolated from human faeces. *Int. J. Syst. Evol. Microbiol.* **50**, 971–978 (2000).
20. Wells, E. & Hylemon, B. Identification and characterization of a bile acid 7 α -dehydroxylation operon in *Clostridium* sp. strain TO-931, a highly active 7 α -dehydroxylating strain isolated from human feces. *Appl. Environ. Microbiol.* **66**, 1107–1113 (2000).
21. Weingarden, A. R. *et al.* Microbiota transplantation restores normal fecal bile acid composition in recurrent *Clostridium difficile* infection. *Am. J. Physiol. Gastrointest. Liver Physiol.* **306**, G310–G319 (2014).
22. Bernstein, H., Bernstein, C., Payne, C. M., Dvorakova, K. & Garewal, H. Bile acids as carcinogens in human gastrointestinal cancers. *Mutat. Res.* **589**, 47–65 (2005).
23. Ng, K. M. *et al.* Microbiota-liberated host sugars facilitate post-antibiotic expansion of enteric pathogens. *Nature* **502**, 96–99 (2013).
24. Jarchum, I., Liu, M., Shi, C., Equinda, M. & Pamer, E. G. Critical role for MyD88-mediated neutrophil recruitment during *Clostridium difficile* colitis. *Infect. Immun.* **80**, 2989–2996 (2012).
25. Jarchum, I., Liu, M., Lipuma, L. & Pamer, E. G. Toll-like receptor 5 stimulation protects mice from acute *Clostridium difficile* colitis. *Infect. Immun.* **79**, 1498–1503 (2011).
26. Rea, M. C. *et al.* Thuricin CD, a posttranslationally modified bacteriocin with a narrow spectrum of activity against *Clostridium difficile*. *Proc. Natl Acad. Sci. USA* **107**, 9352–9357 (2010).

Acknowledgements E.G.P. received funding from US National Institutes of Health (NIH) grants R01 AI42135 and AI95706, and from the Tow Foundation. J.B.X. received funding from the NIH Office of the Director (DP2OD008440), NCI (U54 CA148967), and from a seed grant from the Lucille Castori Center for Microbes, Inflammation, and Cancer. C.G.B. was supported by a Medical Scientist Training Program grant from the National Institute of General Medical Sciences of the NIH (award number T32GM07739, awarded to the Weill Cornell/Rockefeller/Sloan-Kettering Tri-Institutional MD-PhD Program).

Author Contributions C.G.B. and E.G.P. designed the experiments and wrote the manuscript with input from co-authors. C.G.B. performed animal experiments and most analyses. V.B., R.R.S., J.B.X., C.S. and C.G.B. performed microbiota time-series inference modelling and analysis. P.T.M. and C.G.B. designed and performed *ex vivo* experiments. L.L., A.G., A.V. D.N. and M.K. performed 16S amplicon quantification and multiplexed sequencing (454, MiSeq) and contributed to data analysis. M.R.M.v.d.B., R.R.J., Y.T., E.L., C.G.B. and E.G.P. assessed clinical parameters and supervised patient cohort analysis. N.C.T. and C.G.B. performed metagenomic shotgun sequencing analysis. J.R.C. and H.L. developed the metabolomics analysis platform and performed quantification of bile acid species.

Author Information Study sequence data are deposited in the National Center for Biotechnology Information Sequence Read Archive under accession number SRP045811. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to E.G.P. (pamere@mskcc.org).

METHODS

Mouse husbandry. All experiments were performed with C57BL/6J female mice, 6–8 weeks old, purchased from Jackson Laboratories and housed in sterile cages with irradiated food and acidified water. Mouse handling and weekly cage changes were performed by investigators wearing sterile gowns, masks, and gloves in a sterile biosafety hood. All animals were maintained in a specific-pathogen-free facility at Memorial Sloan-Kettering Cancer Center Animal Resource Center. After co-housing for at least 2 weeks, animals (individuals or colonies, as indicated per experiment) were separately housed and randomly assigned to experimental groups. For experiments involving *C. difficile* infection, mice were administered 1,000 *C. difficile* VPI 10463 spores in PBS by oral gavage. All animal experiments were performed at least twice unless otherwise noted. Experiments were performed in compliance with Memorial Sloan-Kettering Cancer Center institutional guidelines and approved by the institution's Institutional Animal Care and Use Committee.

Murine *C. difficile* susceptibility time-course experiments. Mice from three separately housed colonies were kept in the same facility and administered clindamycin (administered by intraperitoneal injection, 200 µg daily), ampicillin (administered in drinking water, 0.1 g l⁻¹), or enrofloxacin (administered in drinking water, 0.4 g l⁻¹) for 3 days (days -2 to 0). At each time point after antibiotic cessation (days 1, 6, 10, 14, and 21), one mouse from each of the three colonies was randomly selected to be single-housed, infected with *C. difficile*, and analysed, yielding triplicate biological measurements per group, per time point. Intestinal content (faeces) was sampled before infection challenge for multiparallel 16S amplicon sequencing and microbiota structure analysis. *C. difficile* susceptibility was determined by selective culture and enumeration of c.f.u. from intestinal content (caecum and colon) 24 h after challenge.

Murine *in vivo* adoptive transfer experiments. Six colonies of mice ($n = 30$ total) were administered antibiotics as described previously²⁷ and subsequently individually housed and assigned randomly to one of three groups. Two days after antibiotics, groups of individually housed mice ($n = 10$ per group) received either 1,000,000 c.f.u. of a four-bacteria suspension (containing equal numbers of *C. scindens* (ATCC35704), *Barnesiella intestihominis* (isolated from murine faeces in-house), *Pseudoflavonifractor capillosus* (ATCC29799), and *Blautia hansenii* (ATCC27752)), a suspension containing 1,000,000 *C. scindens*, or vehicle (PBS) by gavage. All bacteria were grown under anaerobic conditions in reduced brain-heart infusion media supplemented with yeast extract and cysteine except for *B. intestihominis*, which was grown in liquid Wilkins-Chalgren media, and re-suspended in anaerobic PBS before administration to animals. Adoptive transfers of the suspensions were performed once daily for 2 consecutive days before challenge with *C. difficile* VPI 10463 (1,000 spores in PBS). *C. difficile* bacteria and cytotoxin were quantified in faecal samples obtained from mice 24 h after infection challenge. Animals were monitored for 21 days after infection challenge and weight loss was recorded.

Murine *ex vivo* adoptive transfer experiments. Three individually housed mice were administered 200 µg of clindamycin by intraperitoneal injection and killed 24 h later. Intestinal content was harvested from the ilea of killed animals, immediately transferred to an anaerobic chamber, and re-suspended in anaerobic PBS. Fractions containing 100 mg of intestinal content from each mouse were distributed and received either *C. scindens* (100,000 c.f.u.) or vehicle (anaerobic PBS). A third fraction was pre-treated with cholestyramine (1.5 mg) before receiving *C. scindens*. After transfer, the each suspension was inoculated with vegetative *C. difficile* (200 c.f.u.), incubated at 37 °C for 60 h, and *C. difficile* bacteria were quantified by overnight culture on selective media.

Quantitative *C. difficile* culture and toxin A and B. The quantities of *C. difficile* c.f.u. and cytotoxin in the intestinal (caecal) contents of animals were determined as described previously².

Enzymatic assay of secondary bile acid abundance. The relative abundances of primary and secondary bile acids in the intestinal content of killed animals was quantified using an enzymatic assay as described previously²⁸.

Sample collection and DNA extraction. Intestinal microbiota content samples were obtained, snap-frozen, stored, and DNA extracted as described previously²⁹. Briefly, a frozen aliquot (~100 mg) of each sample was suspended, while frozen, in a solution containing 500 µl of extraction buffer (200 mM Tris, pH 8.0/200 mM NaCl/20 mM EDTA), 200 µl of 20% SDS, 500 µl of phenol:chloroform:isoamyl alcohol (24:24:1), and 500 µl of 0.1-mm diameter zirconia/silica beads (BioSpec Products). Microbial cells were lysed by mechanical disruption with a bead beater (BioSpec Products) for 2 min, after which two rounds of phenol:chloroform:isoamyl alcohol extraction were performed. DNA was precipitated with ethanol and re-suspended in 50 µl of TE buffer with 100 µg ml⁻¹ RNase. The isolated DNA was subjected to additional purification with QIAamp Mini Spin Columns (Qiagen). Specimen collection from patients and analysis of the biospecimen group was approved by the Memorial Sloan-Kettering Cancer Center Institutional Review Board. All participants provided written consent for specimen collection and analysis.

Quantification 16S copy number density by rtPCR. DNA extracted from intestinal content samples (faeces) was subjected to rtPCR of 16S rRNA using 0.2 µM concentrations of the broad-range bacterial 16S primers 517F (5'-GCCAGCAG CCGCGGTAA-3') and 798R (5'-AGGGTATCTAATCCT-3') and the DyNAmo SYBR green rtPCR kit (Finnzymes). Standard curves were generated by serial dilution of the PCR blunt vector (Invitrogen) containing one copy of the 16S rRNA gene derived from a member of the Porphyromonadaceae family. The cycling conditions were as follows: 95 °C for 10 min, followed by 40 cycles of 95 °C for 30 s, 52 °C for 30 s, and 72 °C for 1 min.

Quantification of *baiCD* by PCR. DNA extracted from intestinal content samples (faeces) was subjected to PCR-based detection of the 7α-HSDH-encoding *baiCD* gene as described previously³⁰.

16S rRNA gene amplification and multiparallel sequencing. Amplicons of the V4-V5 16S rRNA region were amplified and sequenced using an Illumina MiSeq platform for samples in the *in vivo* and *ex vivo* adoptive transfer experiments. For each sample, duplicate 50-µl PCR reactions were performed, each containing 50 ng of purified DNA, 0.2 mM dNTPs, 1.5 mM MgCl₂, 1.25 U Platinum Taq DNA polymerase, 2.5 µl of 10× PCR buffer, and 0.2 µM of each primer designed to amplify the V4-V5: 563F (5'-nnnnnnnn-NNNNNNNNNNNN-AYTGGGYDTAAAGN G-3') and 926R (5'-nnnnnnnn-NNNNNNNNNNNN-CCGTCATTTHTTTR AGT-3'). A unique 12-base Golay barcode (Ns) preceded the primers for sample identification³¹, and one to eight additional nucleotides were placed in front of the barcode to offset the sequencing of the primers. Cycling conditions were 94 °C for 3 min, followed by 27 cycles of 94 °C for 50 s, 51 °C for 30 s, and 72 °C for 1 min. A condition of 72 °C for 5 min was used for the final elongation step. Replicate PCRs were pooled, and amplicons were purified using the Qiaquick PCR Purification Kit (Qiagen). PCR products were quantified and pooled at equimolar amounts before Illumina barcodes and adaptors were ligated on using the Illumina TruSeq Sample Preparation protocol. The completed library was sequenced on an Illumina MiSeq platform following the Illumina recommended procedures. Samples in the murine and human *C. difficile* susceptibility time-course experiments were sequenced using the 454 FLX Titanium platform as described previously³². Sequences from allo-HSCT patients were obtained from a previously published study⁸.

Sequence analysis. Sequences were analysed using the mothur³³ (version 1.33.3) pipeline. Potentially chimaeric sequences were removed using the UChime algorithm³⁴. Sequences with a distance-based similarity of 97% or greater were grouped into OTUs using the average neighbour algorithm and classified using the BLAST (megablast) algorithm and the GenBank 16S rRNA reference database; OTU-based microbial diversity was estimated by calculating the Shannon diversity index. Sequence abundance profiles in each sample were used for downstream statistical and modelling analysis. A phylogenetic tree was inferred using Clearcut³⁵, on the 16 s rRNA sequence alignment generated by mothur; unweighted UniFrac⁷ was run using the resulting tree, and principal coordinate analysis was performed on the resulting matrix of distances between each pair of samples. PICRUSt (version 0.9.1)¹⁵ in combination with QIIME (version 1.6.0)³⁶ was used to predict abundances of the gene family responsible for secondary bile acid biosynthesis (KEGG pathway ko00121) for a set of 83 samples. Maximum likelihood phylogenetic trees (Kimura model, bootstrap of 100 replicates) were constructed using the MEGA 6.06 package from representative sequences of intestinal bacteria as described³⁷. Raw sequence data from metagenomic shotgun sequencing of six intestinal (ileal) microbiome samples were pre-processed to remove mouse-derived, duplicate, and low-quality reads as well as low-quality bases in accordance with Human Microbiome Project protocols³⁸. The remaining reads were mapped against a set of proteins associated with the secondary bile acid biosynthesis pathway using RAPSearch version 2.07 (ref. 39). For the subsequent analysis, only hits with an *E* value ≤ 0.1, a minimum alignment length of 30, and a minimum similarity of 50% were considered.

Quantification of secondary bile acid species. Samples of murine intestinal content (faeces, ~30 mg) were homogenized using a handheld homogenizer (Omni International) in 80% aqueous methanol and corrected to a final concentration of 0.5 mg per 10 µl. Samples were then sonicated using a Diagenode sonicator at high power, 6 × 30 s cycles. Four hundred microlitres of this material were removed and 20 µl of internal standard added (25 µM d4-chenodeoxycholic acid in 55%/45% methanol/water (v/v)). A further 1 ml methanol was added to the extract and samples were vortexed at 1,400 r.p.m. for 1 h at 30 °C (Thermomixer, Eppendorf). Remaining solid material was removed by centrifugation (21,000g for 10 min) and the supernatant transferred to a glass tube. A second extraction was performed using 1.5 ml methanol, and combined supernatants were dried under a nitrogen gas stream. Finally samples were re-suspended in 300 µl 55%/45% methanol/water (v/v), filtered through a 3 kDa molecular weight cut-off cartridge (Millipore), and transferred to a mass spectrometry vial containing a reduced volume glass insert. Bile acids were separated using an Agilent 1290 HPLC and Cogent C18 column (2.1 mm × 150 mm, 2.2 µm; MicroSolv Technology). Mobile phase A: water + 0.05% formic acid; mobile phase B: acetone + 0.05% formic acid; flow rate 0.35 ml min⁻¹. Injection volume was

5 µl and the liquid chromatography gradient was from 25% B to 70% B in 25 min. Bile acids were detected using an Agilent 6550 Q-TOF mass spectrometer with JetStream source, operating in negative ionization mode and extended dynamic range. Acquisition was from m/z : 50–1,100 at one spectrum per second; gas temperature: 275 °C; drying gas: 11 l min⁻¹; nebulizer: 30 psig; sheath gas: 325 °C; sheath gas flow 10 l min⁻¹; V_{Cap} 4000 V; fragmentor 365 V and Oct 1 RF 750 V. Bile acids (Extended Data Table 2) were identified by their exact mass and confirmed by chromatographic alignment to authentic standards (purchased from Steraloids or Sigma Aldrich). Abundances of the M-H and M+formate ions were then extracted and summed using ProFinder software (Agilent Technologies) and normalized to the internal standard abundance using Mass Profiler Professional software (Agilent Technologies).

DCA and LCA. *C. difficile* inhibition assays. Growth of *C. difficile* in brain–heart infusion liquid media supplemented with yeast extract and cysteine, with added DCA (0.1%, 0.01%, 0.001%, final concentration, in water vehicle) or LCA (0.01%, 0.001%, final concentration, in 100% ethanol vehicle), or vehicle alone was monitored by attenuation ($D_{600 \text{ nm}}$) using a spectrophotometer.

Statistics. Statistical analyses were performed using the R (v. 3.0.2) and GraphPad Prism (version 6.0c) software packages. The Mann–Whitney rank sum test (two-tailed) was used for comparisons of continuous variables between two groups with similar variances; the Kruskal–Wallis test with Dunn correction for multiple comparisons was used for comparisons of three groups or more ($n \geq 5$ samples per group) with similar variances. In all experiments involving group comparisons, at least six animals were used per group; for these non-parametric tests, it was calculated that a sample size of six per group would be sufficient to detect an effect size of 2 with 90% power ($\alpha = 0.05$)⁴⁰. Data were visualized using bar plots with centre values representing the mean and error bars representing standard error of the mean, and box plots representing the median and interquartile range of the upper and lower quartiles and error bars showing the range. Spearman rank correlation tests (two-tailed) were used to find significant correlations between two continuous variables. After statistical analyses with multiple comparisons, we used the Benjamini–Hochberg method to control the false discovery rate. The log-rank test was used to find significant differences in the survival distributions among *C. difficile*-challenged groups of animals. When possible, investigators were blinded during group allocation and outcome assessment (16S and metagenomic shotgun sequence collection, extraction, quantification, and analysis; microbiota time-series inference modelling; quantification of bile species by enzymatic assay and high-performance liquid chromatography–mass spectrometry; enumeration of *C. difficile* in animal experiments).

Inference modelling of mouse microbiota time-series. To determine the network of bacterial–bacterial interactions and extract native resident bacteria with *C. difficile* inhibitory properties, we applied the Lotka–Volterra dynamics-based framework of ref. 10 to the mouse data set. This inference framework consists of a regularized least-square regression of the observed data points and the known antibiotic signal against the difference of the log-transformed total abundances in time:

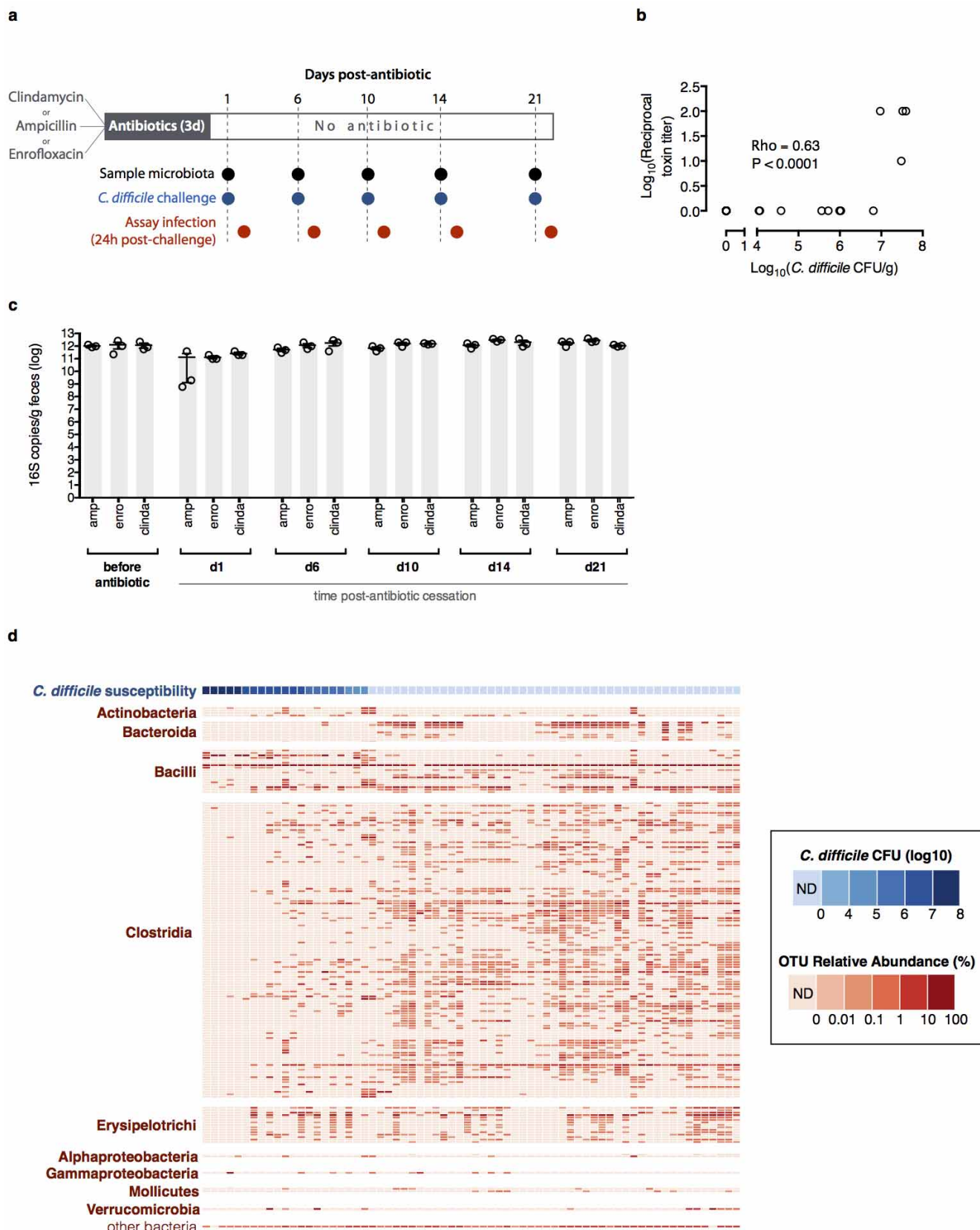
$$\frac{\Delta \ln x_i(t)}{\Delta t} = [\ln(x_i(t + \Delta t)) - \ln(x_i(t))] / \Delta t$$

with $i = 1 \dots N$, where N is the total number of considered OTUs. This results in coefficients characterizing growth, directed interactions, and susceptibilities of each OTU to the external perturbations. The method requires temporal profiles of total abundances of each of the 36 representative OTUs, which were obtained by scaling their normalized abundance from the pyrosequencing run (fraction ranging from 0 to 1) by the total amount of bacteria DNA recovered from each gram of stool or intestinal content. The temporal profile of the *C. difficile* total abundance was obtained from the c.f.u. counts recovered by plating the caecal content after

mouse euthanasia. The last time differences, $\Delta \ln x_i(t_{\text{inoc}})$, were calculated for each mouse as the difference between the total abundance in the intestinal content (faeces) on the day after *C. difficile* inoculation, t_{inoc} , (also the day of mouse euthanasia) minus the total abundance in the content (faeces) before *C. difficile* infection or $\Delta \ln x_i(t_{\text{inoc}}) = \ln x_i^{\text{colon}}(t_{\text{inoc}} + 1) - \ln x_i^{\text{faeces}}(t_{\text{inoc}})$. Similarly the differential profile for *C. difficile* was evaluated from the log-difference in the scaled colony counts for the corresponding faecal and caecal samples. Antibiotic perturbations (ampicillin, clindamycin, or enrofloxacin) were modelled as a discrete signal when administered at day -2 (Fig. 1). The inference algorithm was run on a total of 240 samples and the global model was selected with a threefold cross-validation scheme on the 75 combined time courses to ensure robustness to the introduction of unseen data¹⁰. In particular, the number of data points outnumbers the number of unknowns; that is, the linear system to be solved is overdetermined, ensuring a sufficient number of constraints by the data for inferring the unknown coefficients.

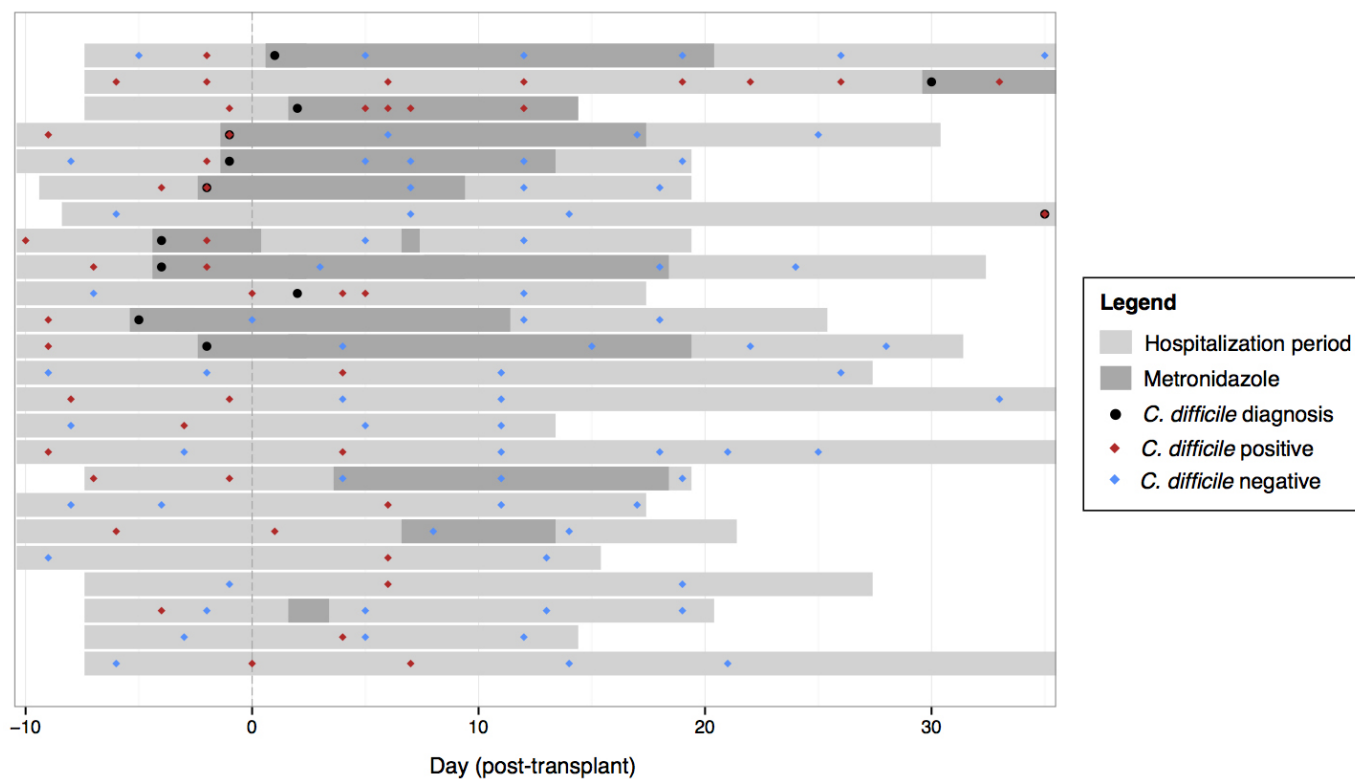
Inference modelling of allo-HSCT patient microbiota time-series. To determine whether commensal–*C. difficile* interactions observed in the mouse data were also conserved in humans, we applied the same inference-modelling framework to data from 24 allo-HSCT hospitalized patients. As above, for each of the 36 OTUs, we determined the log-differential in total abundance as the log-difference of the normalized abundance scaled by the corresponding total bacterial DNA per gram of stool at the next sampling event minus the total abundance at the current sampling time. Differential abundance in *C. difficile* was determined from the rtPCR measurements of the *C. difficile* 16S rRNA gene per gram of faeces. Similarly to the above, we ran the algorithm on a total of 112 samples and the global model was selected applying a threefold cross-validation scheme on the 24 combined time courses. This choice again yields an overdetermined linear system to be solved.

27. Chen, X. *et al.* A mouse model of *Clostridium difficile*-associated disease. *Gastroenterology* **135**, 1984–1992 (2008).
28. Giel, J. L., Sorg, J. A., Sonenshein, A. L. & Zhu, J. Metabolism of bile salts in mice influences spore germination in *Clostridium difficile*. *PLoS ONE* **5**, e8740 (2010).
29. Ubeda, C. *et al.* Vancomycin-resistant *Enterococcus* domination of intestinal microbiota is enabled by antibiotic treatment in mice and precedes bloodstream invasion in humans. *J. Clin. Invest.* **120**, 4332–4341 (2010).
30. Wells, J. E., Williams, K. B., Whitehead, T. R., Heuman, D. M. & Hylemon, P. B. Development and application of a polymerase chain reaction assay for the detection and enumeration of bile acid 7 α -dehydroxylating bacteria in human feces. *Clin. Chim. Acta* **331**, 127–134 (2003).
31. Caporaso, J. G. *et al.* Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. *ISME J.* **6**, 1621–1624 (2012).
32. Ubeda, C. *et al.* Intestinal microbiota containing *Barnesiella* cures vancomycin-resistant *Enterococcus faecium* colonization. *Infect. Immun.* **81**, 965–973 (2013).
33. Schloss, P. D. *et al.* Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Appl. Environ. Microbiol.* **75**, 7537–7541 (2009).
34. Edgar, R. C., Haas, B. J., Clemente, J. C., Quince, C. & Knight, R. UCHIME improves sensitivity and speed of chimera detection. *Bioinformatics* **27**, 2194–2200 (2011).
35. Sheneman, L., Evans, J. & Foster, J. A. Clearcut: a fast implementation of relaxed neighbor joining. *Bioinformatics* **22**, 2823–2824 (2006).
36. Caporaso, J. G. *et al.* QIIME allows analysis of high-throughput community sequencing data. *Nature Methods* **7**, 335–336 (2010).
37. Hall, B. G. Building phylogenetic trees from molecular data with MEGA. *Mol. Biol. Evol.* **30**, 1229–1235 (2013).
38. Human Microbiome Project Consortium Structure. Function and diversity of the healthy human microbiome. *Nature* **486**, 207–214 (2012).
39. Zhao, Y., Tang, H. & Ye, Y. RAPSearch2: a fast and memory-efficient protein similarity search tool for next-generation sequencing data. *Bioinformatics* **28**, 125–126 (2012).
40. Cohen, J. *Statistical Power Analysis for the Behavioral Sciences* (Routledge, 1988).



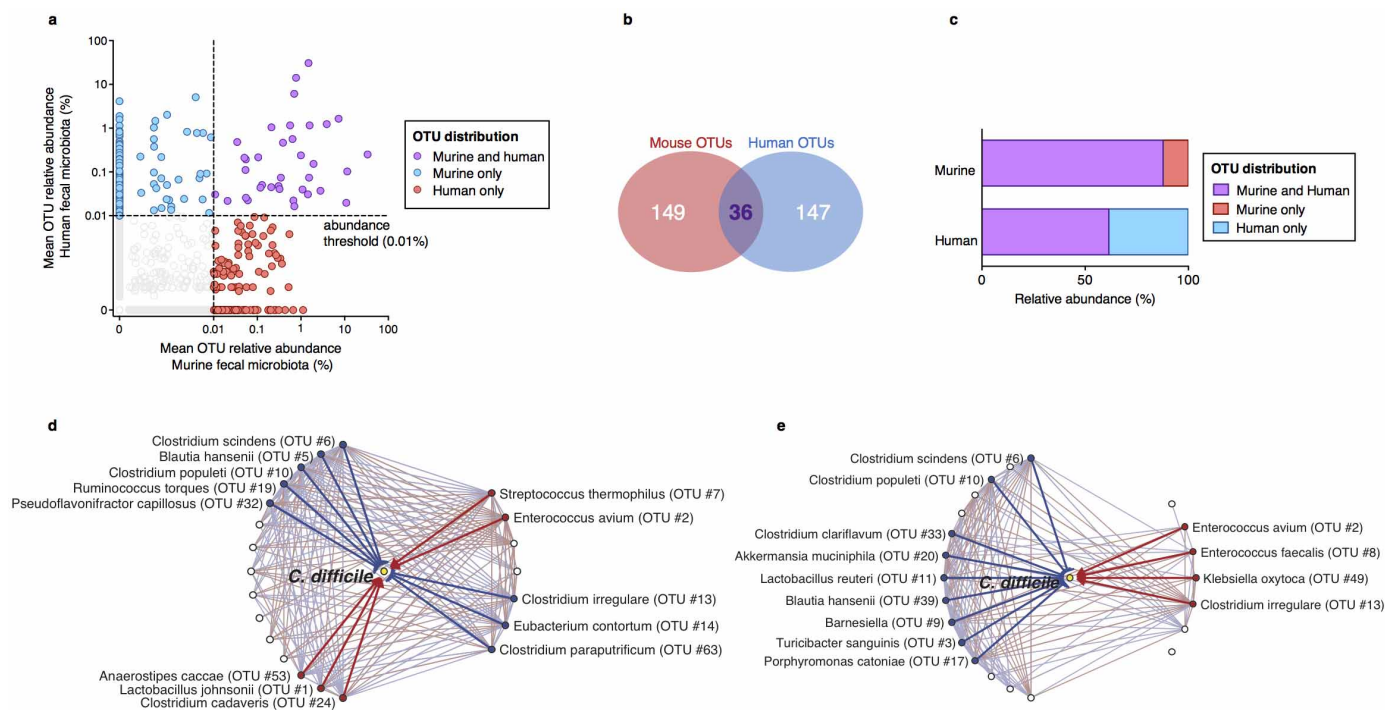
Extended Data Figure 1 | Dynamics of intestinal microbiota structure and *C. difficile* susceptibility after antibiotic exposure. **a**, Strategy for determining *C. difficile* susceptibility duration post-antibiotic exposure ($n=3$ separately-housed mouse colonies per antibiotic arm) and relating infection resistance to microbiota structure. **b**, Correlation of *C. difficile* c.f.u. and toxin in intestinal content following infection. **c**, Intestinal bacterial density of animals

before and after antibiotic exposure. **d**, Relative abundance of bacterial OTUs ($\geq 97\%$ sequence similarity, $>0.01\%$ relative abundance) sorted by class (red) and corresponding *C. difficile* susceptibility (blue) among antibiotic-exposed mice ($n=68$) allowed to recover for variable time intervals prior to *C. difficile* infection challenge. Centre values (mean), error bars (s.e.m.) (c). ND, not detectable.



Extended Data Figure 2 | Allo-HSCT patient timelines and *C. difficile* infection status transitions. Transitions between *C. difficile* (*tcdB*-positive) colonization status in patients receiving allogeneic haematopoietic stem-cell transplantation, as measured by *C. difficile* 16S rRNA abundance during the period of hospitalization (light grey bars). Time points when *C. difficile*

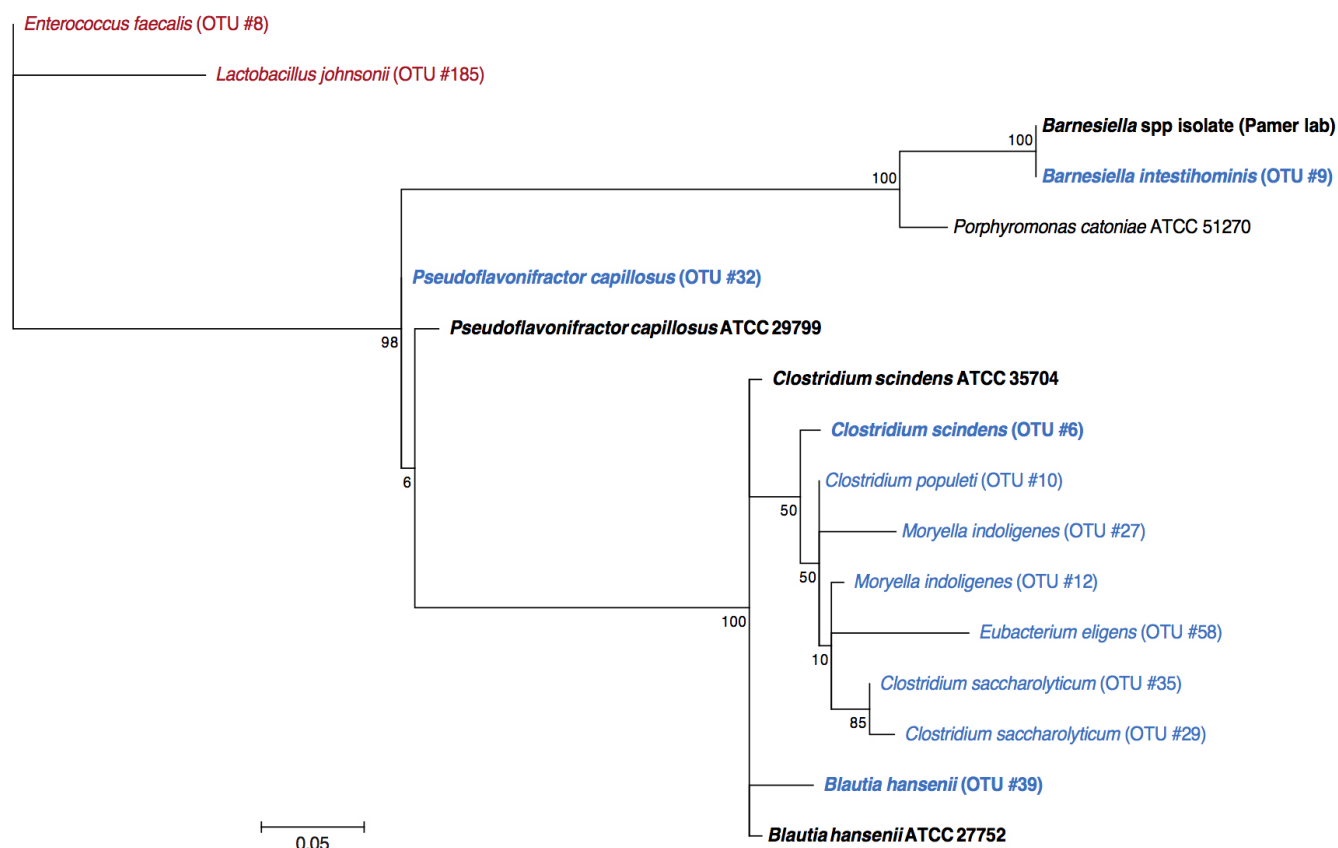
colonization was determined to be positive (red diamonds) and negative (blue diamonds), and when *C. difficile* infection was clinically diagnosed (black dots) and metronidazole was administered (dark grey bars), are displayed relative to the time of transplantation per patient.



Extended Data Figure 3 | Identification of bacteria conserved across human and murine intestinal microbiota predicted to inhibit *C. difficile*.

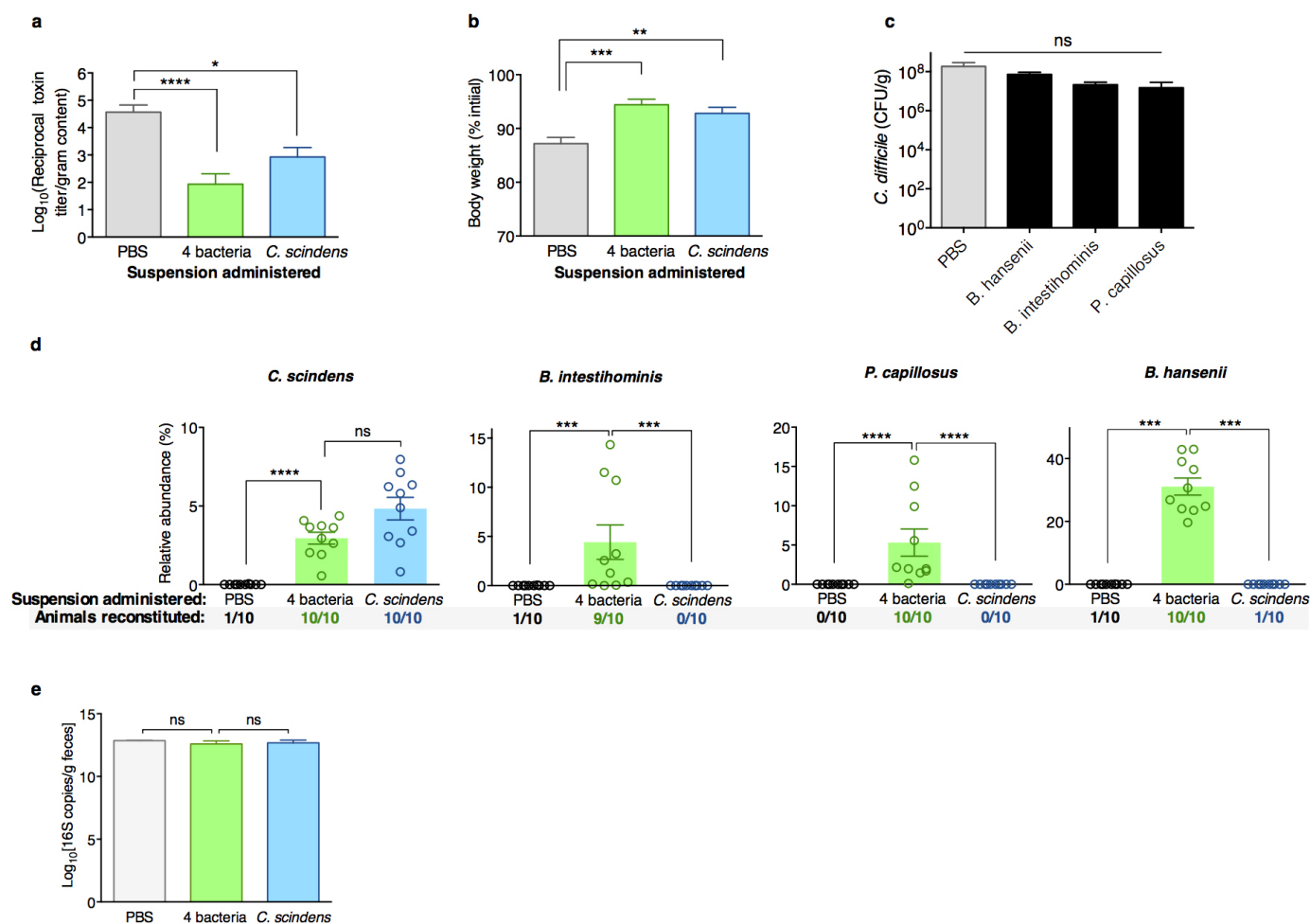
Identification of bacterial OTUs abundant in mice ($n = 68$) and humans ($n = 24$) (a) that account for a minority of OTU membership (b) but the

majority of the structure of the intestinal microbiota of both host species after antibiotic exposure (c). Subnetworks of abundant OTUs predicted inhibit (blue) or positively associate with (red) *C. difficile* in murine (d) and human (e) intestinal microbiota.



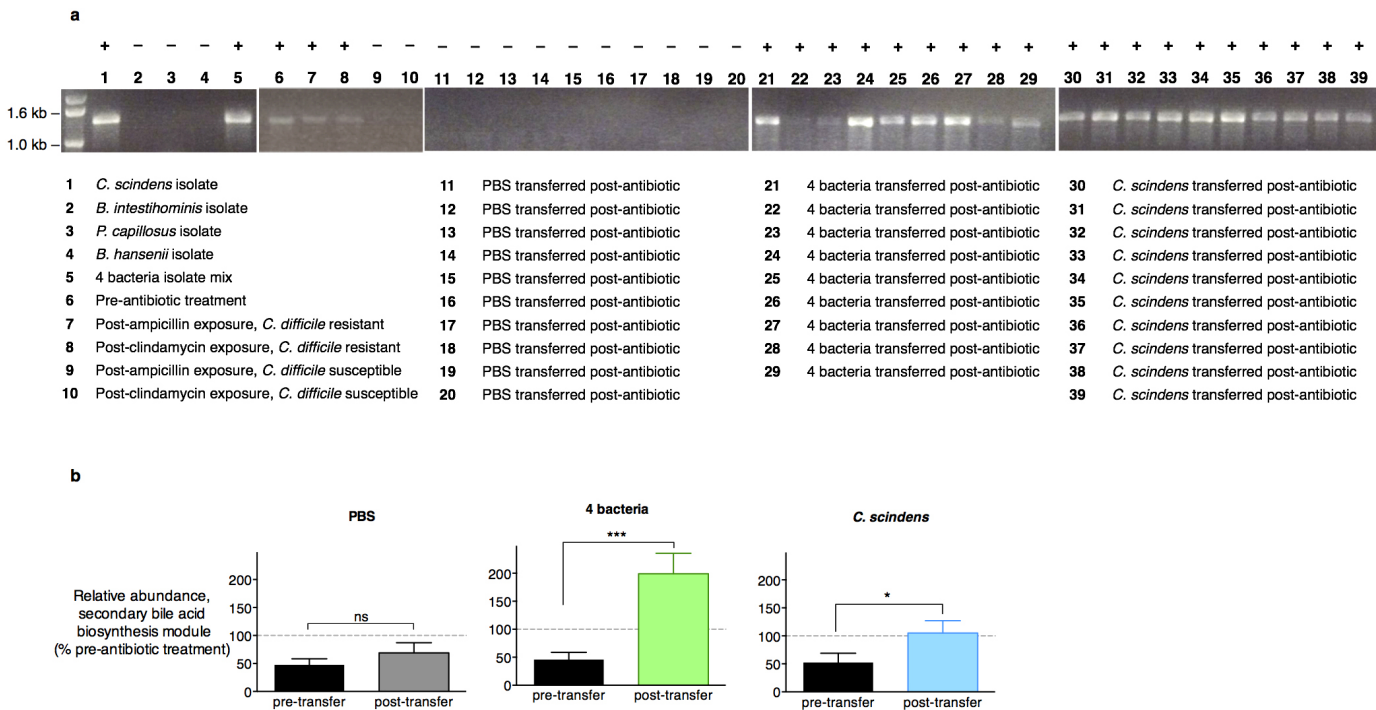
Extended Data Figure 4 | Phylogenetic distribution of resistance-associated intestinal bacteria and isolates selected for adoptive transfer. The maximum likelihood phylogenetic tree (Kimura model, bootstrap of 100 replicates) was constructed using the MEGA 6.06 package from representative sequences of

intestinal bacteria associated with resistance to *C. difficile* infection (blue), including cultured representatives subsequently used in adoptive transfer experiments (bold). The tree was rooted using intestinal bacteria associated with susceptibility to infection (red) as an outgroup.



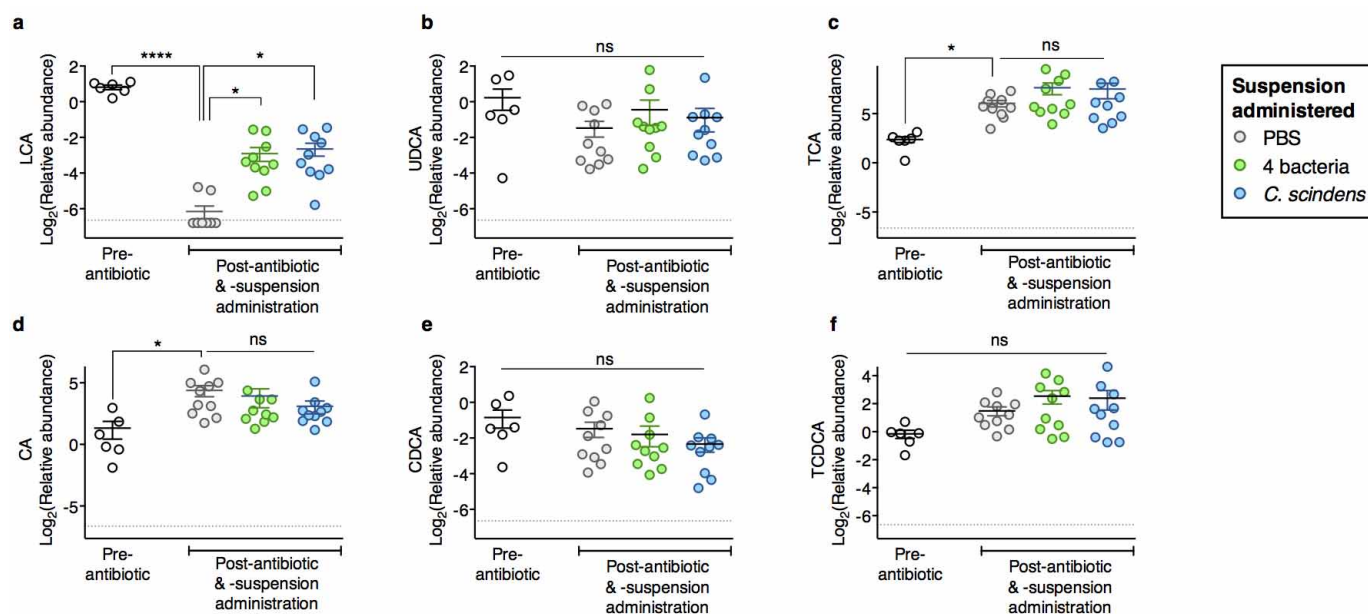
Extended Data Figure 5 | Adoptive transfer and engraftment of four-bacteria consortium or *C. scindens* ameliorates intestinal *C. difficile* cytotoxin load and acute *C. difficile*-associated weight loss. **a**, *C. difficile* toxin load in antibiotic-exposed animals receiving adoptive transfers 24 h after *C. difficile* infection challenge. Animals' weights 48 h after infection challenge and **(b)** *C. difficile* c.f.u. 24 h after infection challenge **(c)**. **d**, Engraftment of bacterial isolates in the intestinal microbiota of antibiotic-exposed animals 2 days after adoptive transfer of *B. intestihominis*, *P. capillosus*, *B. hansenii*, and/or *C. scindens*. **e**, Intestinal bacterial density (faeces) from antibiotic-exposed

mice administered suspensions containing four bacteria, *C. scindens*, or vehicle (PBS) as measured by rtPCR of 16S rRNA genes. **** $P < 0.0001$, *** $P < 0.001$, ** $P < 0.01$, * $P < 0.05$; Mann-Whitney (two-tailed) (**a**, **b**, **d**, **e**), Kruskal-Wallis with Dunn correction **(c)** ($n = 6-10$ per group). Centre values, mean; error bars, s.e.m. Results are representative of at least two independent experiments. Numbers under group columns in **d** denote the number of mice with detectable engraftment of the given bacterium (out of ten possible separately housed animals per group).



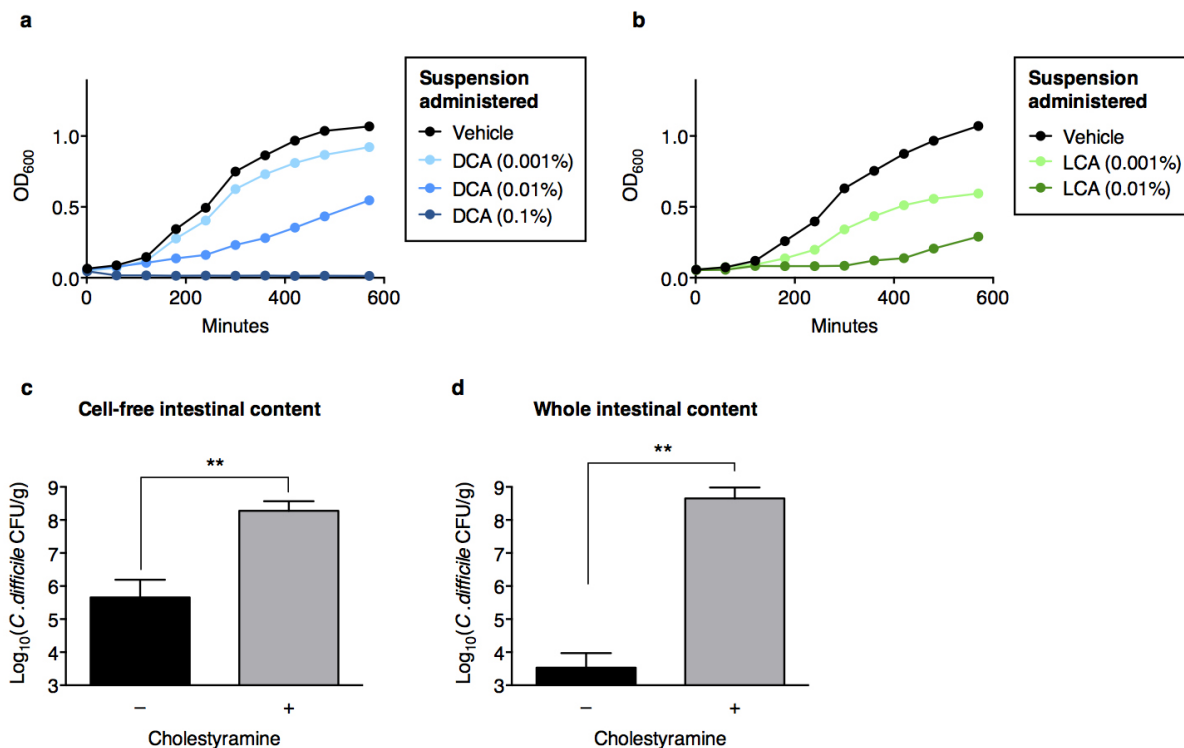
Extended Data Figure 6 | Adoptive transfer of consortia or *C. scindens* restores *baiCD* and the abundance of the gene family responsible for secondary bile acid biosynthesis. **a**, PCR-based detection of the 7 α -HSDH-encoding *baiCD* gene in bacterial isolates, intestinal microbiomes (faeces) of animals before antibiotic exposure, and intestinal microbiomes (faeces) of animals that, after antibiotic exposure, remained susceptible to *C. difficile* or

recovered resistance to infection spontaneously or after adoptive transfer of bacterial isolates. **b**, Reconstituted abundance of the gene family responsible for secondary bile acid biosynthesis, as predicted by PICRUSt, in antibiotic-exposed animals receiving adoptive transfers ($n = 10$ per group). *** $P < 0.001$; * $P < 0.05$; NS, not significant; Mann–Whitney (two-tailed) (**b**). Centre values, mean; error bars, s.e.m.



Extended Data Figure 7 | Impacts of adoptive transfers of bacteria on intestinal abundance of bile acids. Intestinal abundance of the secondary bile acids LCA (a), ursodeoxycholate (UDCA) (b), and primary bile acids (c–f) in

mice after antibiotic exposure and adoptive transfer of bacteria indicated. **** $P < 0.0001$, * $P < 0.05$, NS (not significant); Kruskal–Wallis test with Dunn’s correction. Centre values, mean; error bars, s.e.m.



Extended Data Figure 8 | *C. difficile* growth inhibition by secondary bile acids and intestinal content from antibiotic-naïve animals. Addition of the secondary bile acids DCA (a) or LCA (b) to culture media inhibits *C. difficile*. Bile acid dependent inhibition of *C. difficile* enumerated by recovery of c.f.u.

after inoculation of vegetative *C. difficile* into cell-free (c) or whole (d) intestinal content harvested from C57BL/6J mice ($n = 5$ or 6 per group), with or without pre-incubation with cholestyramine. $**P < 0.01$; Mann-Whitney (two-tailed) (c, d).

Extended Data Table 1 | Characteristics of patients and transplant course

Parameter	No. (% of patients)
Age (years)	
≤29	2/24 (8.3%)
30-39	5/24 (20.8%)
40-49	2/24 (8.3%)
50-59	6/24 (25.0%)
≥60	9/24 (37.5%)
Sex (female)	10/24 (41.7%)
Underlying Disease	
Leukemia	11/24 (45.8%)
Lymphoma	5/24 (20.8%)
Multiple Myeloma	3/24 (12.5%)
Myelodysplastic Syndrome	3/24 (12.5%)
Other	2/24 (8.3%)
Conditioning Intensity	
Nonmyeloablative	4/24 (16.7%)
Reduced intensity	4/24 (16.7%)
Myeloablative	16/24 (66.7%)
T-cell depletion	13/24 (54.2%)
Stem cell source (cord vs. other)	5/24 (20.8%)
Time to engraftment (≥14d) ^{*,†}	5/24 (20.8%)
Fever (T≥100.4) [†]	21/24 (87.5%)
Vital Status: Dead [†]	1/24 (4.2%)
Total	24/24 (100.0%)

* Engraftment was defined as an absolute neutrophil count of more than 500 cells per microlitre for 3 consecutive days.

† Assessed during inpatient allogeneic haematopoietic stem-cell transplantation hospitalization (from 15 days before transplant to 35 days after transplant).

Extended Data Table 2 | Retention times for bile acids quantified by high-performance liquid chromatography–mass spectrometry

Compound	Molecular formula	Accurate mass	Retention time (min)	CAS number
LCA	$C_{24}H_{40}O_3$	376.29775	24.16	434-13-9
UDCA	$C_{24}H_{40}O_4$	392.29266	16.59	128-13-2
DCA	$C_{24}H_{40}O_4$	392.29266	20.42	83-44-3

Blocking PGE₂-induced tumour repopulation abrogates bladder cancer chemoresistance

Antonina V. Kurtova^{1,2}, Jing Xiao¹, Qianxing Mo³, Senthil Pazhanisamy⁴, Ross Krasnow⁴, Seth P. Lerner⁴, Fengju Chen³, Terrence T. Roh^{1,5}, Erica Lay⁴, Philip Levy Ho⁴ & Keith Syson Chan^{1,2,3,4}

Cytotoxic chemotherapy is effective in debulking tumour masses initially; however, in some patients tumours become progressively unresponsive after multiple treatment cycles. Previous studies have demonstrated that cancer stem cells (CSCs) are selectively enriched after chemotherapy through enhanced survival^{1–3}. Here we reveal a new mechanism by which bladder CSCs actively contribute to therapeutic resistance via an unexpected proliferative response to repopulate residual tumours between chemotherapy cycles, using human bladder cancer xenografts. Further analyses demonstrate the recruitment of a quiescent label-retaining pool of CSCs into cell division in response to chemotherapy-induced damages, similar to mobilization of normal stem cells during wound repair^{4–7}. While chemotherapy effectively induces apoptosis, associated prostaglandin E₂ (PGE₂) release paradoxically promotes neighbouring CSC repopulation. This repopulation can be abrogated by a PGE₂-neutralizing antibody and celecoxib drug-mediated blockade of PGE₂ signalling. *In vivo* administration of the cyclooxygenase-2 (COX2) inhibitor celecoxib effectively abolishes a PGE₂- and COX2-mediated wound response gene signature, and attenuates progressive manifestation of chemoresistance in xenograft tumours, including primary xenografts derived from a patient who was resistant to chemotherapy. Collectively, these findings uncover a new underlying mechanism that models the progressive development of clinical chemoresistance, and implicate an adjunctive therapy to enhance chemotherapeutic response of bladder urothelial carcinomas by abrogating early tumour repopulation.

Cytotoxic chemotherapy remains the standard of care for many advanced carcinomas. Although chemotherapy is effective in debulking tumour mass, certain patients show initial response but progressively become unresponsive after multiple treatments. Chemotherapy is administered in cycles to induce fractionated killing of unsynchronized proliferating cancer cells, and treatments are spaced out to allow recovery of normal tissues between cycles⁸. However, repopulation of residual surviving cancer cells also occurs, which is an undesirable phenomenon that limits chemotherapeutic response in subsequent cycles⁸. Recent studies demonstrated that CSCs have a survival advantage in response to chemotherapy^{1–3}. Here we investigate the unexplored concept that CSCs may actively proliferate in response to chemotherapy-induced damages, similar to how tissue resident stem cells mobilize to wound sites during tissue repair^{4–7,9}.

Bladder urothelial carcinomas contain cells that span various cellular differentiation stages^{10–15}, cytokeratin 14 (CK14) marks the most primitive (or least differentiated) cells^{11,13} and patients with abundant CK14 staining correlate with poor survival^{11,13}. Here, comparative analysis of matching pre- and post-chemotherapy patient tissues revealed one group with CK14 staining enrichment/persistence (Fig. 1a and Extended Data Fig. 1a–c) and another group with no CK14 staining after chemotherapy (Fig. 1a and Extended Data Fig. 1a, b, d). Kaplan–Meier analysis

revealed patients with CK14⁺ cancer cell enrichment/persistence showed worse survival (Fig. 1a), justifying further need to investigate their chemotherapeutic response. Using the standard chemotherapy regimen for advanced bladder urothelial carcinomas (that is, gemcitabine and cisplatin (GC)), one chemotherapy cycle effectively reduced the growth rate of all xenograft tumours in comparison to controls (Fig. 1b and Extended Data Fig. 2a), while leading to a generalized enrichment of CK14⁺ cancer cells (1.7–4.3-fold) (Fig. 1c, d and Extended Data Fig. 2b, c). This enrichment is unexpectedly contributed by proliferation marked by mitosis phase protein phospho-histone H3 (Extended Data Fig. 2d, e; white arrows). In addition to the conventional thinking that chemotherapy selects for chemoresistant cancer cells, this active proliferative response may represent a new mechanism contributing to repopulation of residual tumours. To investigate this phenomenon further, we constructed a lentiviral reporter to enable prospective isolation of CK14⁺ cells by fluorescence activated cell sorting (FACS), as CK14 is an intracellular protein that would not allow for cell surface antibody labelling. We subcloned a previously validated gene promoter region of human *KRT14* (ref. 16) into a promoterless lentiviral vector carrying a tdTomato (hK14. tdTomato) red fluorescent protein (Extended Data Fig. 3a). With this reporter stably transduced into urothelial carcinoma cells (Fig. 1e and Extended Data Fig. 3b–d), we could readily detect a tdTomato⁺ (Tm⁺) subpopulation that exclusively expressed CK14 at the protein (Fig. 1f; white arrows) and messenger RNA (Fig. 1g; *KRT14*) levels, while minimally expressing the differentiation marker uropod 1B (Fig. 1g; *UPK1B*). These Tm⁺ CK14⁺ urothelial carcinoma cells represented a subpopulation of our previously reported CD44⁺CD49f⁺ CSCs^{11,13} (Extended Data Fig. 3b–d). Additional functional evaluation validated their unique biological properties being enriched for sphere-forming stem cells *in vitro* (Extended Data Fig. 3e) and tumorigenic cells when engrafted *in vivo* (Extended Data Fig. 3f), thus demonstrating accepted functional criteria for CSCs. To evaluate their chemotherapeutic response, we purified Tm⁺CK14⁺ and Tm[−]CK14[−] cancer cells and evaluated their relative cell viability after GC chemotherapy (Fig. 1h and Extended Data Fig. 4). Tm⁺CK14⁺ cancer cells survived chemotherapy-induced apoptosis significantly better than Tm[−]CK14[−] cells starting at day 3 (Fig. 1h and Extended Data Fig. 4). Concurrent cell cycle analyses revealed an unexpected proliferative response of both subpopulations by entering into S phase at days 2 and 3, respectively (Fig. 1i, j and Extended Data Fig. 5). Interestingly, Tm[−]CK14[−] cancer cells remained proliferative throughout the 11-day time course, whereas Tm⁺CK14⁺ cancer cells gradually returned to a less proliferative state (Fig. 1i, j and Extended Data Fig. 5). Because gemcitabine (a cytidine analogue) and cisplatin preferentially incorporate into proliferating cells to initiate apoptosis via inducing DNA crosslinks, strand breaks and adduct formation¹⁷, we propose that the slower cycling properties of CK14⁺ cells may provide one mechanism by which they evade GC chemotherapy.

¹Department of Molecular & Cellular Biology, Baylor College of Medicine, One Baylor Plaza, Houston, Texas 77030, USA. ²Program in Translational Biology and Molecular Medicine, Baylor College of Medicine, One Baylor Plaza, Houston, Texas 77030, USA. ³Dan L. Duncan Cancer Center and Center for Cell Gene & Therapy, Baylor College of Medicine, One Baylor Plaza, Houston, Texas 77030, USA. ⁴Scott Department of Urology, Baylor College of Medicine, One Baylor Plaza, Houston, Texas 77030, USA. ⁵Summer Medical and Research Training (SMART) Program, Baylor College of Medicine, One Baylor Plaza, Houston, Texas 77030, USA.

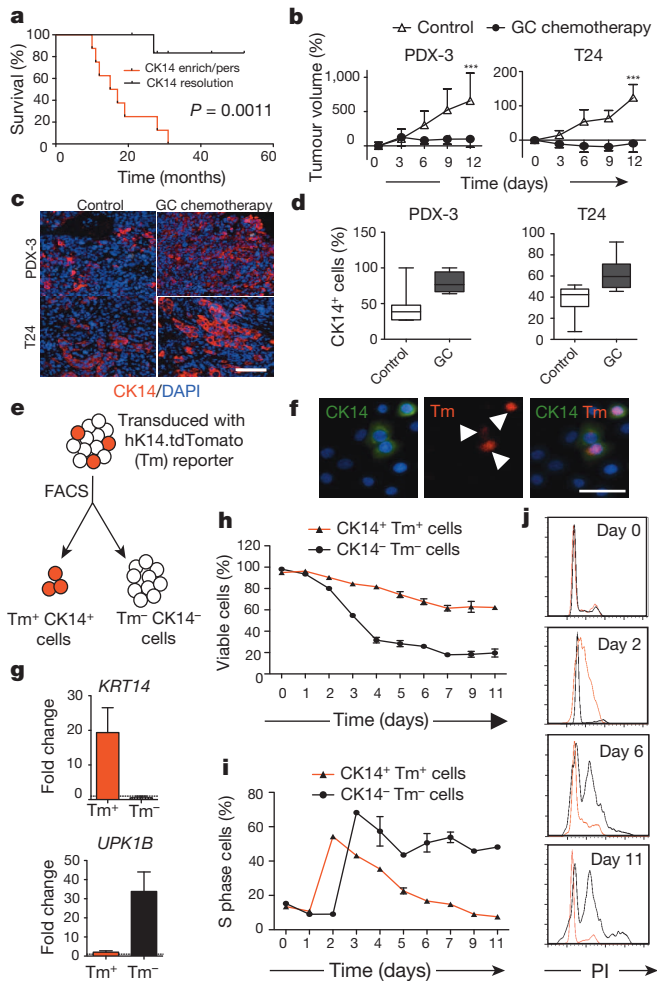


Figure 1 | Cytotoxic chemotherapy induces CK14⁺ cancer cell proliferation despite reducing tumour size. **a**, Kaplan–Meier analysis of bladder-cancer patients with various CK14 staining patterns (enrich/pers denotes an enrichment (increase) or persistence of CK14 staining; resolution denotes an absence) after neoadjuvant chemotherapy ($n = 15$). **b**, Relative change in xenograft tumour size after chemotherapy ($n = 6$ xenograft tumours per group). PDX-3, patient-derived xenograft from patient 3; T24, human bladder cancer cell line-derived xenograft. **c**, **d**, Representative immunofluorescence staining (**c**) and box plots quantifying the percentage of CK14⁺ bladder cancer cells after chemotherapy (**d**). DAPI, 4',6-diamidino-2-phenylindole. **e**, Schematic demonstrating the transduction of hK14.tdTomato lentiviral reporter into urothelial carcinoma cells. **f**, Immunofluorescence staining verifying the specific expression of tdTomato (Tm)-positive signal in CK14⁺ cancer cells. **g**, Quantitative PCR (qPCR) analysis of *KRT14* and *UPK1B* in cancer cell subpopulations in biological duplicates. **h**, **i**, Graphs summarizing viability (**h**) and corresponding percentage of S phase cells (**i**) in two cancer subpopulations after chemotherapy in biological duplicates. **j**, Cell cycle profiles for Tm⁺ CK14⁺ (red line) and Tm⁻ CK14⁻ (black line) cancer cells at indicated days after chemotherapy. PI, propidium iodide. Data in **b** represent mean \pm s.e.m.; box plots in **c** show twenty-fifth to seventy-fifth percentiles, with line indicating the median, and whiskers indicating the smallest and largest values; data in **g–i** show mean and range. *** $P < 0.001$ (log-rank test (**a**) and two-tailed Student's *t*-test (**b**)). Scale bars, 100 μ m.

To explore this hypothesis *in vivo*, we used the classical pulse-chase labelling approach to mark quiescent cells^{18,19}, followed by chemotherapy treatment to analyse their proliferative response (Fig. 2a). We repetitively pulsed xenograft tumours with the thymidine analogue 5-iodo-2'-deoxyuridine (IdU), followed by a 4–8-week chase period, during which dividing cells diluted out IdU whereas quiescent cells retained IdU¹⁸ (Fig. 2a). Effective IdU pulsing was validated (Extended Data Fig. 6a) and a quiescent cancer subpopulation (1.08–5.48%) was consistently found (Extended Data Fig. 6a–d), which we term label-retaining

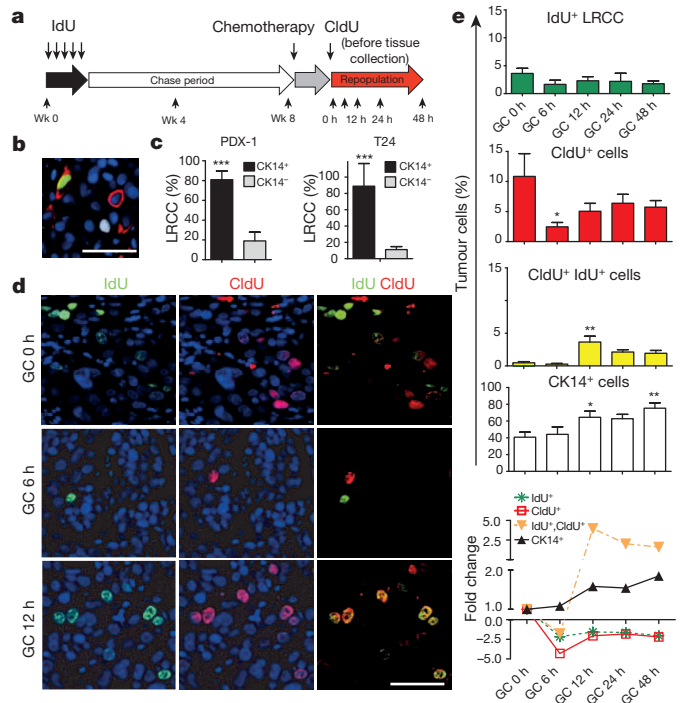


Figure 2 | Chemotherapy recruits CK14⁺ LRCCs to proliferate and divide. **a**, Experimental approach to label and localize LRCCs in bladder xenografts. Wk, week. **b**, Immunofluorescence staining to locate CK14⁺ cancer cells (red) that co-localize with LRCCs (green IdU⁺ cells) or proliferating cells (white CldU⁺ cells) at 8 weeks after IdU chase. **c**, Quantification of LRCCs within cancer subpopulations ($n = 4$ xenografted tumours). **d**, Immunofluorescence staining to evaluate the percentage and localization of LRCCs (green, IdU⁺), proliferating cells (red, CldU⁺) and dividing LRCCs (yellow, IdU⁺ CldU⁺) at various time after chemotherapy ($n = 4$ xenografted tumours). **e**, Time kinetics revealing the percentage of LRCCs (green, IdU⁺), proliferating cells (red, CldU⁺), dividing LRCCs (yellow, IdU⁺ CldU⁺), and CK14⁺ cancer cells (white) at various time after chemotherapy ($n = 4$ xenografted tumours). All data represent mean \pm s.e.m. * $P < 0.05$, ** $P < 0.01$ (one-way analysis of variance (ANOVA) followed by Dunnett's test for multiple comparisons). Scale bars, 50 μ m.

cancer cells (LRCCs) (green cells in Fig. 2b, Extended Data Fig. 6a, b). These IdU⁺ LRCCs predominately localized within CK14⁺ cells (Fig. 2b, c; red) and were mutually exclusive to proliferating cells labelled by 5-chloro-2'-deoxyuridine (CldU) (Fig. 2b; white) or other proliferation markers (Extended Data Fig. 6e–g). Notably, chemotherapy significantly reduced the fraction of CldU⁺ proliferating cells while not altering IdU⁺ LRCCs initially (Fig. 2d, e; 6 h). These results confirmed our *in vitro* findings that gemcitabine and cisplatin preferentially induce apoptosis of dividing cells¹⁷ (Fig. 1h, i). This is paradoxically followed by a significant increase in IdU⁺ LRCCs that were simultaneously incorporated with CldU (Fig. 2d, e; 12 h). These IdU⁺ CldU⁺ cells constantly appeared as a two-cell state immediate adjacent to each other (Fig. 2d; yellow cells), suggestive of LRCC recruitment to divide after chemotherapy-induced apoptotic damage. This trend coincided with an increase in CldU⁺ proliferating and CK14⁺ cancer cells (Fig. 2e and Extended Data Fig. 6h; 12 h), which gradually decreased from 24 h and onwards, corresponding with a plateau of single CldU⁺ proliferating cancer cells (Fig. 2e). Collectively, these results illustrate that quiescent LRCCs are unexpectedly recruited into cell division after chemotherapy treatment, contributing to repopulation of residual tumours.

Next, we explored the molecular mechanisms underlying tumour repopulation. After chemotherapy, it is known that soluble factors are released from apoptotic cells to attract immune cells for clearance²⁰. Paradoxically, some of these factors are also known to stimulate epithelial cell proliferation during wound repair. One example is PGE₂, which was demonstrated to promote stem-cell expansion²¹ and mobilization²².

Here, we investigate PGE₂ and CSCs in the context of chemotherapy, which was not studied previously. Urothelial carcinoma cells were exposed to chemotherapy *in vitro* (Fig. 3a) and their viability was quantified by flow cytometry analyses of mitochondrial and plasma membrane permeability (Fig. 3b) and by cleaved caspase-3 (CASP3) (Fig. 3c). The time kinetics of chemotherapy-induced apoptosis (Fig. 3b, c) is closely associated with an increase in the expression of COX2 (the enzyme mediating PGE₂ production) and PGE₂ release (Fig. 3d). We further validated these findings *in vivo*, by analysing the co-localization of PGE₂, CASP3 and CK14⁺ cancer cells, and their associated changes after chemotherapy (Fig. 3e, f). Similar to *in vitro*, chemotherapy induced apoptosis as measured by a CASP3-positive stain, which peaked at 6 h and gradually reduced thereafter (Fig. 3e, f), which was closely followed by PGE₂ release from 6 to 24 h after chemotherapy (Fig. 3f). Furthermore, PGE₂ localization (Fig. 3e; red) was often found adjacent to CK14⁺

cancer cells (Fig. 3e; green). Altogether, these findings demonstrated that chemotherapy-induced apoptosis is closely followed by the release of PGE₂, which preceded that of tumour repopulation (Fig. 3g).

To establish a biological role for PGE₂ in promoting tumour repopulation, we investigated how exogenous dmPGE₂ (a stabilized PGE₂ analogue) could modulate sphere-forming stem cells *in vitro*²³. Exogenous dmPGE₂ treatment significantly increased the frequency of sphere-forming cells (Fig. 3h). To evaluate PGE₂ in the context of chemotherapy, we applied PGE₂-containing supernatant obtained from chemotherapy-treated cells to urothelial carcinoma cells (Fig. 3a), and compared their sphere-forming ability to those receiving exogenous dmPGE₂ (Fig. 3i). Although both conditions significantly induced sphere-forming cells, supernatant from chemotherapy-treated cells induced a higher frequency (Fig. 3i). Notably, functional blockade of PGE₂ by a neutralizing antibody significantly reduced the ability of factor-containing supernatant to generate sphere-forming cells (Fig. 3i). Furthermore, the addition of the pharmacological COX2 inhibitor celecoxib, alone or in combination with a PGE₂ neutralizing antibody, further reduced their frequencies (Fig. 3i).

Because more than 70% of patients with advanced bladder cancer do not benefit from neoadjuvant chemotherapy^{24–26}, and celecoxib is a US Food and Drug Administration (FDA)-approved drug, we evaluated whether the attenuation of PGE₂ signalling *in vivo* could enhance chemotherapeutic response (Fig. 4a). Although chemotherapy is clinically administered in multiple cycles, few preclinical studies administered chemotherapy beyond one cycle or a single dose. To bridge this gap, we recapitulated the clinical regimen (Fig. 4a–e) in advanced bladder cancer xenografts, including one derived from a patient who was resistant to chemotherapy (Fig. 4e and Extended Data Fig. 1a). During the first chemotherapy cycle, both PDX-3 and T24 xenograft lines were responsive to chemotherapy (GC) and combination therapy (celecoxib plus GC) by reducing tumour size (Fig. 4b, e; days 3–12). Nevertheless, celecoxib combination treatment did abrogate notable expansion of CK14⁺ cancer cells induced by chemotherapy (Extended Data Fig. 7a–c). Remarkably, manifestation of chemoresistance occurred progressively after successive chemotherapy cycles (Fig. 4b, e; black circle), and celecoxib combination treatment (celecoxib plus GC) enhanced chemotherapeutic response more evidently in later treatment cycles (Fig. 4b, e; asterisk). T24 bladder cancer xenografts were limited to two chemotherapy cycles owing to distal lung metastases, and celecoxib combination treatment significantly diminished metastatic lung foci (Fig. 4c, d). Because celecoxib is known to suppress prostacyclin and predisposes patients to thrombosis²⁷, patients with cancer prescribed celecoxib as adjuvant therapy will probably be co-administered with low-dose aspirin for cardioprotection. We confirmed the antithrombotic efficacy of low-dose aspirin without affecting the overall efficacy of celecoxib (Extended Data Fig. 8).

Here, our findings highlight a new mechanism by which CK14⁺ CSCs contribute to the progressive development of chemoresistance, via their proliferative response to fuel repopulation of residual tumours during the gap periods between chemotherapy cycles (Extended Data Fig. 10). Global gene expression profiling further revealed a parallel to tissue wound response, that chemoresistant tumours elicited enrichment of a ‘wound-response gene signature’ (Fig. 4f, g and Extended Data Fig. 9a, b), which could be abrogated by combination celecoxib treatment (Fig. 4h and Extended Data Fig. 9c). The *PTGS2* gene, which encodes COX2 and is a core pathway component for PGE₂ signalling, is among these wound-response genes (Fig. 4g, h). Similar findings were found in human bladder carcinomas that are resistant to chemotherapy (Fig. 4f and Extended Data Fig. 9a, d, e), thus confirming the importance and generalization of PGE₂/COX2 signalling during development of chemoresistance, while other metabolites of the arachidonic acid pathway may also be involved. By validating the efficacy of celecoxib in abrogating chemoresistance in xenografts derived from a patient who was resistant to chemotherapy, this preclinical finding provides considerable implications for immediate clinical translation, as the standard of care

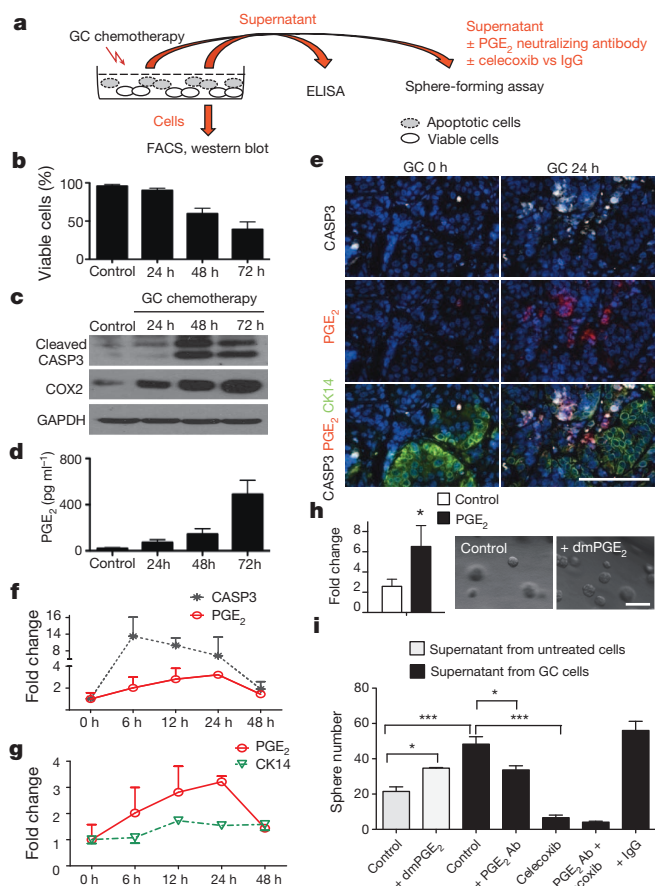


Figure 3 | PGE₂ release from neighbouring apoptotic cells induces CSC repopulation. **a**, Experimental approach to evaluate the role of PGE₂ release in modulating sphere-forming CSCs. **b**, FACS quantification of the percentage of viable cells at various time after chemotherapy ($n = 3$ experiments). **c**, Western blot analysing the correlative protein level of cleaved caspase-3 (CASP3) and COX2 at various time after chemotherapy (representative blot from $n = 3$). GAPDH was a loading control. **d**, Temporal release of PGE₂ by chemotherapy-treated cancer cells, measured by ELISA ($n = 3$ experiments). **e**, Immunofluorescence staining demonstrating relative localization of CASP3-positive apoptotic cells (white), PGE₂-positive signals (red) and CK14-positive cancer cells (green) in chemotherapy- or vehicle-treated xenograft tumours *in vivo* ($n = 4$ xenografted tumours). **f**, **g**, Relative fold changes in CASP3-positive (black), PGE₂-positive (red) and CK14-positive (green) cancer cells with time. **h**, Exogenous effects of dmPGE₂ on sphere-forming CSCs ($n = 3$ experiments). **i**, Inhibitory effects of a PGE₂-neutralizing antibody (Ab) and celecoxib in modulating sphere-forming CSCs ($n = 4$ biological replicates). All data represent mean \pm s.e.m. * $P < 0.05$, *** $P < 0.001$ (Student's *t*-test (**h**) and one-way ANOVA followed by Dunnett's test for multiple comparisons (**i**)). Scale bars, 100 μ m.

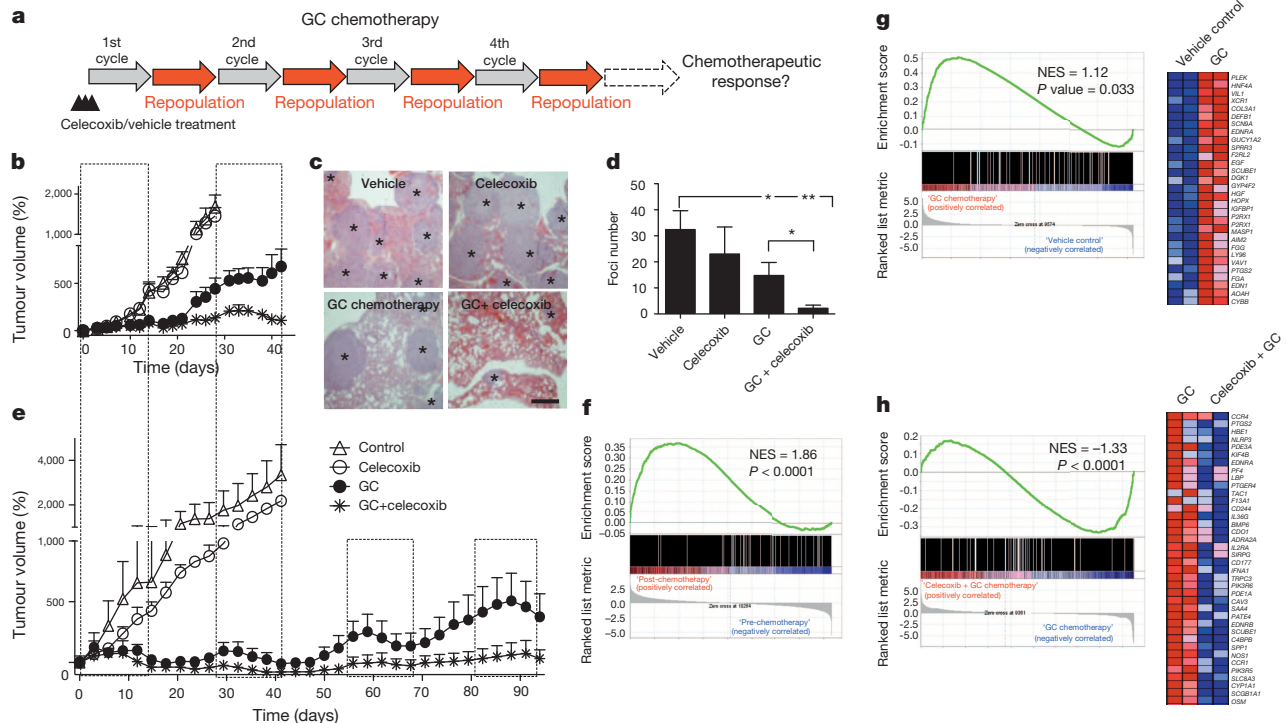


Figure 4 | Celecoxib drug-mediated inhibition of PGE₂ pathway abrogates progressive development of chemoresistance. **a**, *In vivo* preclinical chemotherapy recapitulates clinical regimen with multiple treatment cycles and gap periods. **b**, Percentage change in tumour size after chemotherapy, in the presence or absence of celecoxib ($n = 12$ xenografted tumours (T24) per group). **c**, Representative haematoxylin and eosin images showing lung metastatic foci (asterisks) from various treatments. Scale bar, 500 μ m. **d**, Quantification of lung metastatic foci across various treatments ($n = 5$ mice per group). **e**, Percentage change in tumour size of patient xenografts (PDX-3) after chemotherapy, in the presence or absence of celecoxib ($n = 6$ xenografted tumours per group). The four dashed boxes indicate the time frame

of each GC chemotherapy cycle. **f**, Gene set enrichment analysis (GSEA) validated enrichment of 'wound-response gene signature' after chemotherapy in patients with bladder cancer classified as non-responders. NES, normalized enrichment score. **g**, GSEA revealing enrichment of wound-response gene signature in chemotherapy-treated bladder cancer xenografts. Heat map demonstrating genes within leading edge, including *PTGS2* (encoding COX2), a core pathway component for PGE₂. **h**, GSEA and heat map highlighting abrogation of wound-response gene signature after celecoxib combination treatment. All data represent mean \pm s.e.m. Tumour volume measurements are relative to mean tumour volume at day 0. * $P < 0.05$, ** $P < 0.01$ (one-way ANOVA followed by Dunnett's test for multiple comparisons).

for advanced bladder cancers is limited to cytotoxic chemotherapy and surgery with no other therapeutics^{28–30}. This approach to abrogate chemoresistance may extend to other epithelial tumour types on further functional validation.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 2 June 2013; accepted 4 November 2014.

Published online 3 December 2014.

- Visvader, J. E. & Lindeman, G. J. Cancer stem cells in solid tumours: accumulating evidence and unresolved questions. *Nature Rev. Cancer* **8**, 755–768 (2008).
- Clevers, H. The cancer stem cell: premises, promises and challenges. *Nature Med.* **17**, 313–319 (2011).
- Kreso, A. & Dick, J. E. Evolution of the cancer stem cell model. *Cell Stem Cell* **14**, 275–291 (2014).
- Beachy, P. A., Karhadkar, S. S. & Berman, D. M. Tissue repair and stem cell renewal in carcinogenesis. *Nature* **432**, 324–331 (2004).
- Arwert, E. N., Hoste, E. & Watt, F. M. Epithelial stem cells, wound healing and cancer. *Nature Rev. Cancer* **12**, 170–180 (2012).
- Ito, M. *et al.* Stem cells in the hair follicle bulge contribute to wound repair but not to homeostasis of the epidermis. *Nature Med.* **11**, 1351–1354 (2005).
- Mascré, G. *et al.* Distinct contribution of stem and progenitor cells to epidermal maintenance. *Nature* **489**, 257–262 (2012).
- Kim, J. J. & Tannock, I. F. Repopulation of cancer cells during therapy: an important cause of treatment failure. *Nature Rev. Cancer* **5**, 516–525 (2005).
- Wilson, A. *et al.* Hematopoietic stem cells reversibly switch from dormancy to self-renewal during homeostasis and repair. *Cell* **135**, 1118–1129 (2008).
- Chan, K. S. *et al.* Identification, molecular characterization, clinical prognosis, and therapeutic targeting of human bladder tumor-initiating cells. *Proc. Natl Acad. Sci. USA* **106**, 14016–14021 (2009).
- Volkmer, J. P. *et al.* Three differentiation states risk-stratify bladder cancer into distinct subtypes. *Proc. Natl Acad. Sci. USA* **109**, 2078–2083 (2012).

- Ho, P. L., Lay, E. J., Jian, W., Parra, D. & Chan, K. S. Stat3 activation in urothelial stem cells leads to direct progression to invasive bladder cancer. *Cancer Res.* **72**, 3135–3142 (2012).
- Ho, P. L., Kurtova, A. & Chan, K. S. Normal and neoplastic urothelial stem cells: getting to the root of the problem. *Nature reviews. Urology* **9**, 583–594 (2012).
- Shin, K. *et al.* Cellular origin of bladder neoplasia and tissue dynamics of its progression to invasive carcinoma. *Nature Cell Biol.* **16**, 469–478 (2014).
- Van Batavia, J. *et al.* Bladder cancers arise from distinct urothelial sub-populations. *Nature Cell Biol.* **16**, 982–991 (2014).
- Wang, X., Zinkel, S., Polonsky, K. & Fuchs, E. Transgenic studies with a keratin promoter-driven growth hormone transgene: prospects for gene therapy. *Proc. Natl Acad. Sci. USA* **94**, 219–226 (1997).
- Kelland, L. The resurgence of platinum-based cancer chemotherapy. *Nature Rev. Cancer* **7**, 573–584 (2007).
- Cotsarelis, G., Sun, T. T. & Lavker, R. M. Label-retaining cells reside in the bulge area of pilosebaceous unit: implications for follicular stem cells, hair cycle, and skin carcinogenesis. *Cell* **61**, 1329–1337 (1990).
- Chen, J. *et al.* A restricted cell population propagates glioblastoma growth after chemotherapy. *Nature* **488**, 522–526 (2012).
- Green, D. R. The end and after: how dying cells impact the living organism. *Immunity* **35**, 441–444 (2011).
- Goessling, W. *et al.* Genetic interaction of PGE2 and Wnt signaling regulates developmental specification of stem cells and regeneration. *Cell* **136**, 1136–1147 (2009).
- Hoggatt, J. *et al.* Differential stem- and progenitor-cell trafficking by prostaglandin E2. *Nature* **495**, 365–369 (2013).
- Lukacs, R. U., Goldstein, A. S., Lawson, D. A., Cheng, D. & Witte, O. N. Isolation, cultivation and characterization of adult murine prostate stem cells. *Nature Protocols* **5**, 702–713 (2010).
- Meeks, J. J. *et al.* A systematic review of neoadjuvant and adjuvant chemotherapy for muscle-invasive bladder cancer. *Eur. Urol.* **62**, 523–533 (2012).
- Sternberg, C. N. *et al.* ICUD-EAU International Consultation on Bladder Cancer 2012: chemotherapy for urothelial carcinoma-neoadjuvant and adjuvant settings. *Eur. Urol.* **63**, 58–66 (2013).
- Niegijsch, G. *et al.* Neoadjuvant chemotherapy in patients with muscle-invasive bladder cancer: which patients benefit? *Eur. Urol.* **64**, 355–357 (2013).

27. Coxib and traditional NSAID Trialists' (CNT) Collaboration. Vascular and upper gastrointestinal effects of non-steroidal anti-inflammatory drugs: meta-analyses of individual participant data from randomised trials. *Lancet* **382**, 769–779 (2013).
28. Wu, X. R. Urothelial tumorigenesis: a tale of divergent pathways. *Nature Rev. Cancer* **5**, 713–725 (2005).
29. Flaig, T. W. & Theodore, D. Bladder cancer in 2011: the dawn of personalized medicine. *Nature reviews. Urology* **9**, 65–66 (2012).
30. Knowles, M. A. & Hurst, C. D. Molecular biology of bladder cancer: new insights into pathogenesis and clinical diversity. *Nature Rev. Cancer* <http://dx.doi.org/10.1038/nrc3817> (in the press).

Acknowledgements We would like to thank S. Tsai, J. Rosen, L. Donehower, M. Brenner and B. Deneen for their editorial suggestions, and grant funding from the National Cancer Institute CA129640, CA175397, V Scholar Award, Dan L Duncan Career Award, and Bladder Cancer Partnership (K.S.C.), CPRIT pre-doctoral fellowship RP101499 (A.V.K.), AUA Research Scholar Award (P.L.H.) and CPRIT training grant RP140102 for SMART program (T.T.R.). This work is supported by the Genomic & RNA Profiling Core (L. White), Cytometry & Cell Sorting Core (NIAID AI036211, NCI CA125123 and RR024574; J. Sederstorm), and Human Tissue Acquisition & Pathology Core

(NCI CA125123; M. Ittmann) at Baylor College of Medicine, the Texas Advanced Computing Center at UT Austin and Rice University for providing High Performance Computing resources. We dedicate this work to the memory of J. Pride (a patient with bladder cancer and friend of I. Weissman), who funded the original work of K.S.C., which served as an important foundation leading to the current study.

Author Contributions K.S.C. and A.V.K. conceived, designed and performed data analysis. A.V.K., J.X. and S.P. performed the functional xenograft tumour experiments. A.V.K., J.X., E.L. and P.L.H. performed molecular analyses experiments. A.V.K. and J.X. performed, and Q.M. and F.C. analysed, the RNA-seq studies. S.P.L. provided human samples. R.K. coordinated human sample acquisition, R.K., P.L.H. and T.T.R. performed immunohistochemical and pathological analysis. K.S.C. and A.V.K. wrote the manuscript.

Author Information RNA-seq raw data have been deposited in the NCBI BioProject database under accession number PRJNA263046. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to K.S.C. (kc1@bcm.edu).

METHODS

Acquisition of human bladder urothelial carcinoma tissues. With approval from Institutional Review Board (H-26809) and informed consents, we obtained matching samples of formalin-fixed paraffin-embedded human bladder tumours before (collected at the time of transurethral resection of bladder tumour) and after (collected at the time of radical cystectomy) chemotherapy. Our pathologist selected representative specimens on review of haematoxylin and eosin slides.

Generation of patient-derived xenografts from human urothelial carcinomas. All mice (immunocompromised *Rag2*^{-/-} *Il2rg*^{-/-} double-knockout mice; Taconic) were housed in the Animal Facility at Baylor College of Medicine. Freshly isolated patient cancer tissues (clinical information is summarized in Extended Data Fig. 1a) were obtained with patient consent under Institutional Review Board (H-25099) approval at Baylor College of Medicine. Tumour tissues were subdivided into 2–5 mm³ pieces, coated with growth-factor-reduced matrigel (BD) and media in a 1:1 ratio, and embedded within the subcutaneous space underneath the skin of mice (male or female, 6–8 weeks old). Successfully xenografted tumours were then enzymatically dissociated into single cell suspensions and directly passaged *in vivo* without prior culturing *in vitro*, as previously reported^{10,11}. Animal work was performed in accordance to ethical regulations approved by Baylor College of Medicine Institutional Animal Care and Use Committee (IACUC; AN-5803).

In vitro treatment. Immortalized bladder cancer cell line T24 (high-grade) was purchased from ATCC and maintained in DMEM (HyClone), supplemented with 10% FBS (Gibco) and penicillin-streptomycin (Cellgro). Cells used were authenticated and tested for mycoplasma at the Tissue and Cell Culture Core Laboratory at Baylor College of Medicine. Cells were incubated with gemcitabine (7.5 ng ml⁻¹; Sigma) 2 h before addition of cisplatin (100 µg ml⁻¹; Sigma). Determination of cell viability after treatment was based on the analysis of mitochondrial transmembrane potential by 3,3'-dihexyloxycarbocyanine iodide (DiOC₆; Molecular Probes, Invitrogen) and cell membrane permeability to propidium iodide (Sigma) by FACS (BD Fortessa) at the indicated time points (24, 48 and 72 h). CASP3 cleavage was measured at the indicated time points using standard immunoblotting with rabbit-polyclonal CASP3 antibody (Cell Signaling, 9662; 1:1,000), COX2 expression was measured with rabbit-polyclonal COX2 antibody (Cell Signaling, 4842; 1:1,000), GAPDH expression was measured with mouse-monoclonal GAPDH antibody (Santa Cruz Biotechnology, sc-137179; 1:400). Appropriate secondary antibodies (anti-rabbit/mouse IgG, horseradish peroxidase (HRP)-linked antibody, Cell Signaling; 1:2,000) were used for detection.

In vivo treatment. When xenograft tumours reached a palpable size, mice (male and female) were assigned into different experimental groups: vehicle control (12.5% dimethylsulphoxide (DMSO) in saline) group; GC chemotherapy treatment group; celecoxib treatment group; and celecoxib plus GC combinatory treatment group. According to the Institutional Animal Care and Use Committees (IUCUC) protocol (AN-5803), the maximal allowable size of the tumours could not exceed 10% of mouse body weight. For systemic administration of a single GC treatment, mice were treated with either vehicle or sequential intraperitoneal injection of gemcitabine (60 mg kg⁻¹) and cisplatin (6 mg kg⁻¹) given 4 h apart. For one GC cycle, cisplatin (6 mg kg⁻¹) was only applied after the first gemcitabine (40 mg kg⁻¹) treatment on day 2, followed by three consecutive treatments of gemcitabine on days 5, 8 and 11 (GC→G→G→G). The tumours were collected 24 h after the last treatment. In celecoxib combination treatment experiments, mice were pretreated with celecoxib (5 mg kg⁻¹ daily; Sigma PZ0008) via intraperitoneal injection 2 days before the initiation of chemotherapy, followed by daily intraperitoneal administration of celecoxib during chemotherapy or vehicle treatment cycle.

Immunohistochemical analysis of CK14 in patient samples. Immunohistochemical staining was performed using keratin 14 polyclonal antibody (Covance, PRB-155P; 1:1,000). Every sample was scored for its CK14 staining by determining the percentage of visual fields that were CK14 negative (0%), CK14 low-infiltrating (<25%), CK14 high-infiltrating (25–100%), and CK14 basal-restricted (staining restricted to the basal layer of urothelium). The final result for each case was the average score of all visual fields. COX2 staining was performed on serial sections using COX2 monoclonal antibody (Cayman Chemical, 160112; 1:100).

Immunofluorescence analyses. Tumour sections were analysed following standard haematoxylin and eosin procedures or immunostaining protocols as previously reported¹⁰. Nikon microscopy and NIS element software were used for imaging and quantification. Primary antibodies used are listed as follows: 5-bromo-2'-deoxyuridine (BrdU) (for IdU, BD 347580; 1:100), BrdU (BD 347583; 1:100), BrdU (for CldU, Abcam, ab6326; 1:100), CK14 (Covance, PRB-155P; 1:1,000), CK14 (Abcam, ab77684; 1:100), cleaved CASP3 (Cell Signaling, 9661; 1:400), PGE₂ (Abcam, ab2318; 1:100), and phospho-histone H3 (Cell Signaling, 3377; 1:1,000).

Generation of hK14.tdTomato lentiviral reporter construct. We developed a lentiviral construct that reports on the activity of human *KRT14* promoter by subcloning a previously characterized and validated promoter region of human *KRT14* (ref. 16) (obtained from ATCC, MBA-124) using EcoRI and BamHI restriction

enzymatic digestion, and inserting this promoter fragment into a promoterless lentiviral vector (Clontech, 631754) tagged with a DD-tdTomato (Tm) red fluorescent protein (Clontech Lenti-X DD-tdTomato vector). DD-Tm protein is rapidly targeted to proteasome degradation unless incubated with Shield1 (Clontech) *in vitro*.

Limiting dilution analysis of tumour subpopulations. For *in vivo* limiting dilution assay, Tm⁺ CK14⁺ and Tm⁻ CK14⁻ tumour cells were purified by FACS (BD FACS Aria Sorter) and serially diluted to the appropriate cell dose. Cells were injected and the number of tumours formed from each cell dose injected was scored. The frequency of cancer stem cells had been calculated using the ELDA software (<http://bioinf.wehi.edu.au/software/elda/index.html>) provided by the Walter and Eliza Hall Institute.

qRT-PCR. RNA isolation (RNAspin Mini, Illustra, GE Healthcare) and reverse transcription (Superscript III First-Strand Synthesis, Invitrogen) were performed according to the manufacturer's instructions. We used Taqman Gene Expression assay mix (Applied Biosystems) and FAM-labelled Taqman probes to detect expression of the following genes: *KRT14* (Applied Biosystems, Hs00265033_m1) and *UPK1B* (Applied Biosystems, Hs00199583_m1). Relative qPCR ($\Delta\Delta C_t$ method) was performed in triplicate using ABI Prism 7000 Sequence Detection System (Applied Biosystems). Gene expression levels were normalized to *GAPDH* transcript levels (Applied Biosystems, Hs99999905_m1).

Cell cycle analysis. Tm⁺ and Tm⁻ cells subjected to *in vitro* GC chemotherapy were collected using trypsin (Gibco), washed twice with PBS and fixed with 70% ethanol for 30 min at room temperature. Cells were stained with propidium iodide (25 µg ml⁻¹ solution in 4 mM citrate buffer supplemented with RNase, 10% Triton X-100 and PEG 6000; all from Sigma) and analysed by FACS (gating strategy excluded cell doublets from analysis).

Dual labelling of slow-cycling and active-proliferating cancer cells. Mice were labelled by intraperitoneal injection of IdU (Sigma-Aldrich; 50 mg kg⁻¹ bodyweight), BrdU (Sigma-Aldrich; 50 mg kg⁻¹ bodyweight) or CldU (Sigma-Aldrich). For labelling of slow-cycling cells, repetitive pulsing of IdU (50 mg kg⁻¹ bodyweight) was given for 5 consecutive days, followed by various long chase periods (4–8 weeks) to localize LRCCs. For labelling of proliferating cells, CldU was given 1 h before tumour tissue collection to localize active cycling cancer cells.

PGE₂ ELISA assay. To measure PGE₂ release, bladder cancer cells T24 were cultured in DMEM supplemented with 2% FBS and treated with gemcitabine (7.5 ng ml⁻¹; Sigma) and cisplatin (100 µg ml⁻¹; Sigma). Supernatants were collected 24, 48 and 72 h after treatment initiation. PGE₂ levels in the supernatants were measured by ELISA kit (R&D Systems).

dmPGE₂ exogenous treatment. T24 cancer cells were plated in a semi-solid sphere-forming assay in the presence of 0.38 ng ml⁻¹ dmPGE₂, 100 nM of neutralizing PGE₂ antibody (Cayman Chemical, 10009814), 5 µM celecoxib (Sigma) or vehicle for 12 days²³. Clonal derivation of spheres from single cells was ensured and monitored as described²³.

Thromboprophylaxis with aspirin. Immunocompromised mice bearing T24 xenografted tumours were treated with celecoxib and GC chemotherapy in the presence or absence of aspirin (Sigma) (37.9 µg g⁻¹ twice a week, equivalent of 162.5 mg daily dose in human) for two cycles, and antithrombotic activity of aspirin was tested by analysing tail bleeding time. In brief, tails of anaesthetized mice (1–3% isoflurane in 100% oxygen at a flow rate of 1–2 L min⁻¹) were transected 1 mm from the tip with a scalpel, the remaining tail was immersed in 37 °C saline, and the time until bleeding stopped for more than 1 min was observed and recorded.

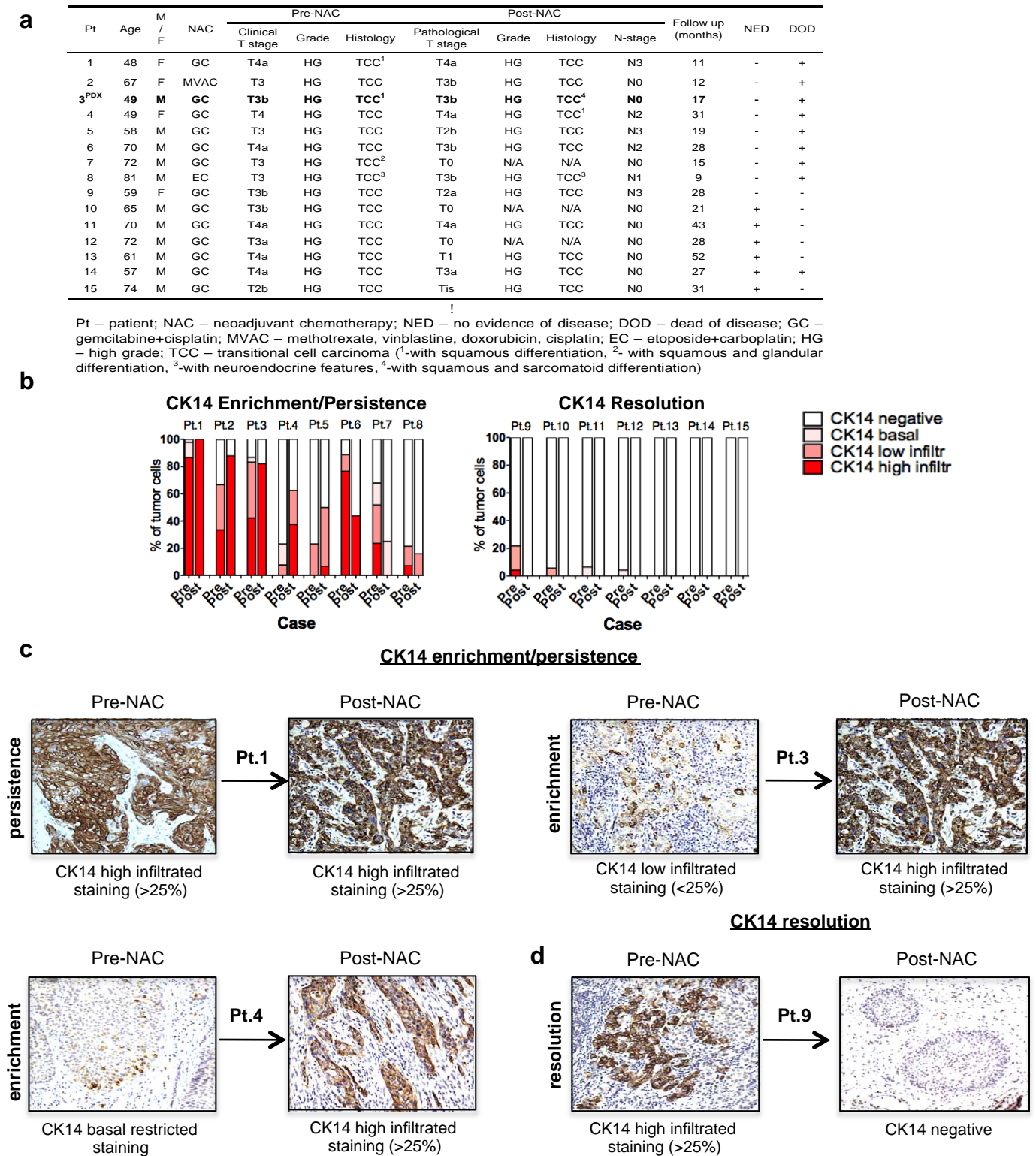
RNA-sequencing. Tumour tissues were collected from the animals and snap-frozen in liquid nitrogen. RNA isolation was performed using Illustra RNAspin mini kit (GE Healthcare) upon manufacturer instructions. The Genomic and RNA Profiling Core at Baylor College of Medicine conducted sample quality checks using the NanoDrop spectrophotometer and Agilent Bioanalyzer 2100. A double-stranded DNA library was created using 1 µg of total RNA, preparing the fragments for hybridization onto a flowcell. First, cDNA was created using the fragmented 3' poly(A) selected portion of total RNA and random primers. Libraries were created from the cDNA by first blunt ending the fragments, attaching an adenosine to the 3' end and finally ligating unique adapters to the ends. The ligated products were then amplified using 15 cycles of PCR. The resulting libraries were quantitated using the NanoDrop spectrophotometer and fragment size assessed with the Agilent Bioanalyzer. A qPCR quantitation was performed on the libraries to determine the concentration of adaptor ligated fragments using a Bio-Rad iCycler iQ Real-Time PCR Detection System and a KAPA Library Quant Kit. Using the concentration from the Bio-Rad qPCR machine, 18 pM of library was loaded onto a flowcell and amplified by bridge amplification using the Illumina cBot machine. A paired-end 100 cycle run was used to sequence the flowcell on the HiSeq 2500 Sequencing System.

Bioinformatics analysis. The Philadelphia neoadjuvant chemotherapy (NAC) data (GSE48277) were obtained from Gene Expression Omnibus (<http://www.ncbi.nlm.nih.gov/geo>), which consisted of microarray gene expression data of urothelial tumours before and after NAC treatment. The microarray data were quantile normalized

and then \log_2 transformed. If there were multiple probes for a gene, we used the probe with median expression value for GSEA. For the RNA-seq data analysis, the reads were trimmed from the both ends to 90 nucleotides, and then mapped to the human genome (UCSC hg19) using Tophat with NCBI RefSeq genes as the reference and up to two possible mismatches. To reduce possible biases caused by RT-PCR, read duplicates were removed using samtools. HTseq (<http://www-huber.embl.de/users/anders/HTSeq>) was used to determine the number of reads falling in the known genes. The counts that indicate gene expression values were normalized to have the same total counts for each sample. The normalized counts were \log_2 transformed and 1 was added to read counts to avoid \log_2 transformation of zeros. Genes with small expression value (count < 6) for all samples were not used for GSEA. The wound response gene signature (GO:0009611—response to wounding) was obtained from Amigo (<http://amigo.geneontology.org>) consisting of 1,163 genes.

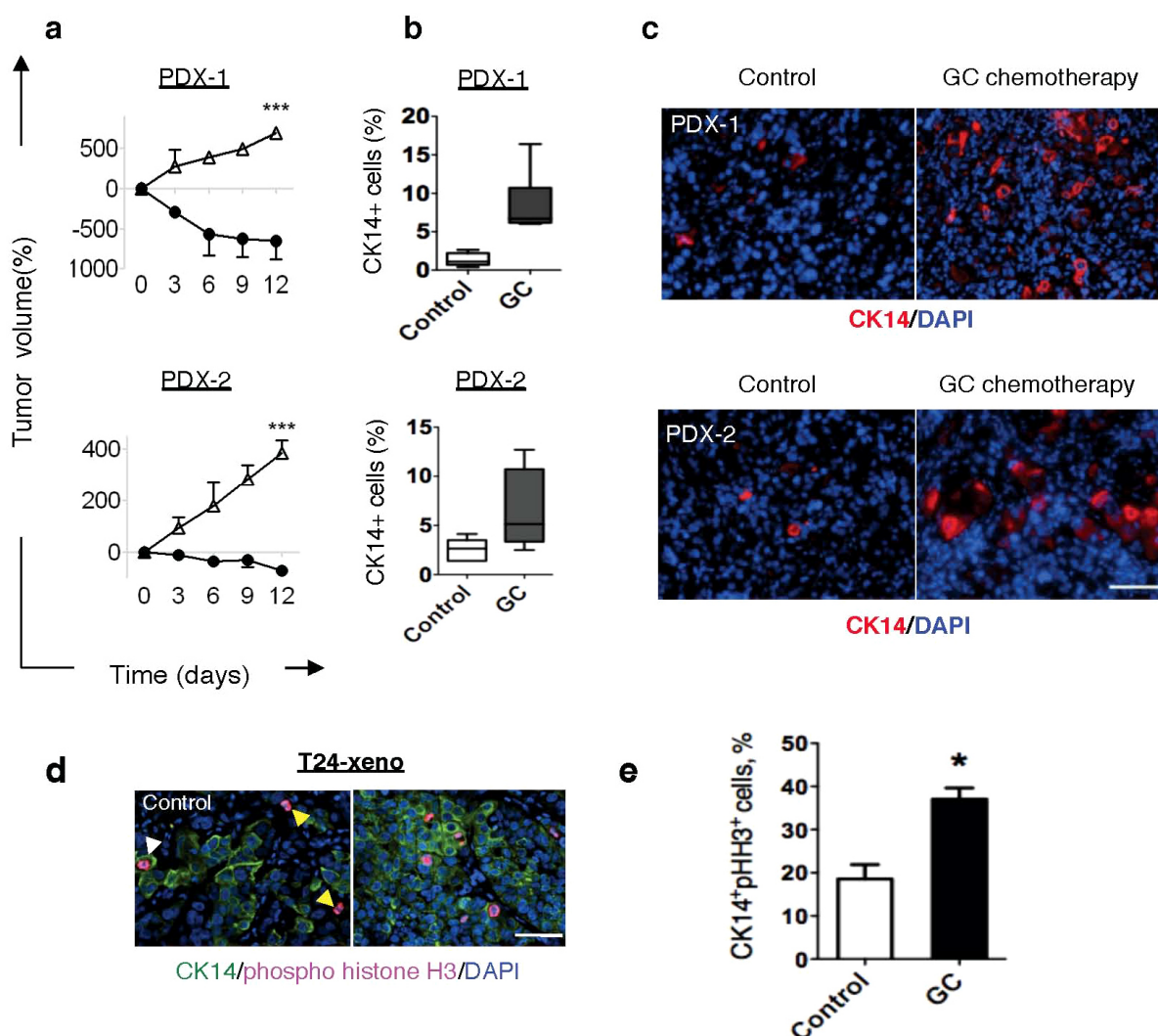
Statistical analysis. All statistical analyses were performed using GraphPad Prism (GraphPad Software) and included two-tailed Student's *t*-test or one-way ANOVA

followed by Dunnett's test for comparing the means of two or multiple groups, respectively. $P < 0.05$ was considered statistically significant. No formal randomization was used in any experiment; for *in vivo* experiments animals were unbiasedly assigned into different treatment groups. Group allocation and outcome assessment was not performed in a blinded manner. No animals or samples were excluded from data analysis. On the basis of our preliminary results, we estimated that the tumour sizes in the GC treatment group are about 200 mm³ and we expected celecoxib plus GC combination treatment would reduce the tumour size to 80–100 mm³. With an estimated standard deviation of 50 mm³ within group, we expected to have at least 80% power to detect differences of tumour volume ($\alpha = 0.05$, two-tailed *t*-test) using a sample size of 6 mice in each group. For *in vivo* experiments performed with patient-derived tumour cells, the sample size was dependent on the number of viable cells isolated upon tumour dissociation. Generally, at least six tumours were analysed for each group tested. In each set of data analysis, estimate variation is indicated in each figure as s.e.m. No statistical method was used to predetermine sample size.



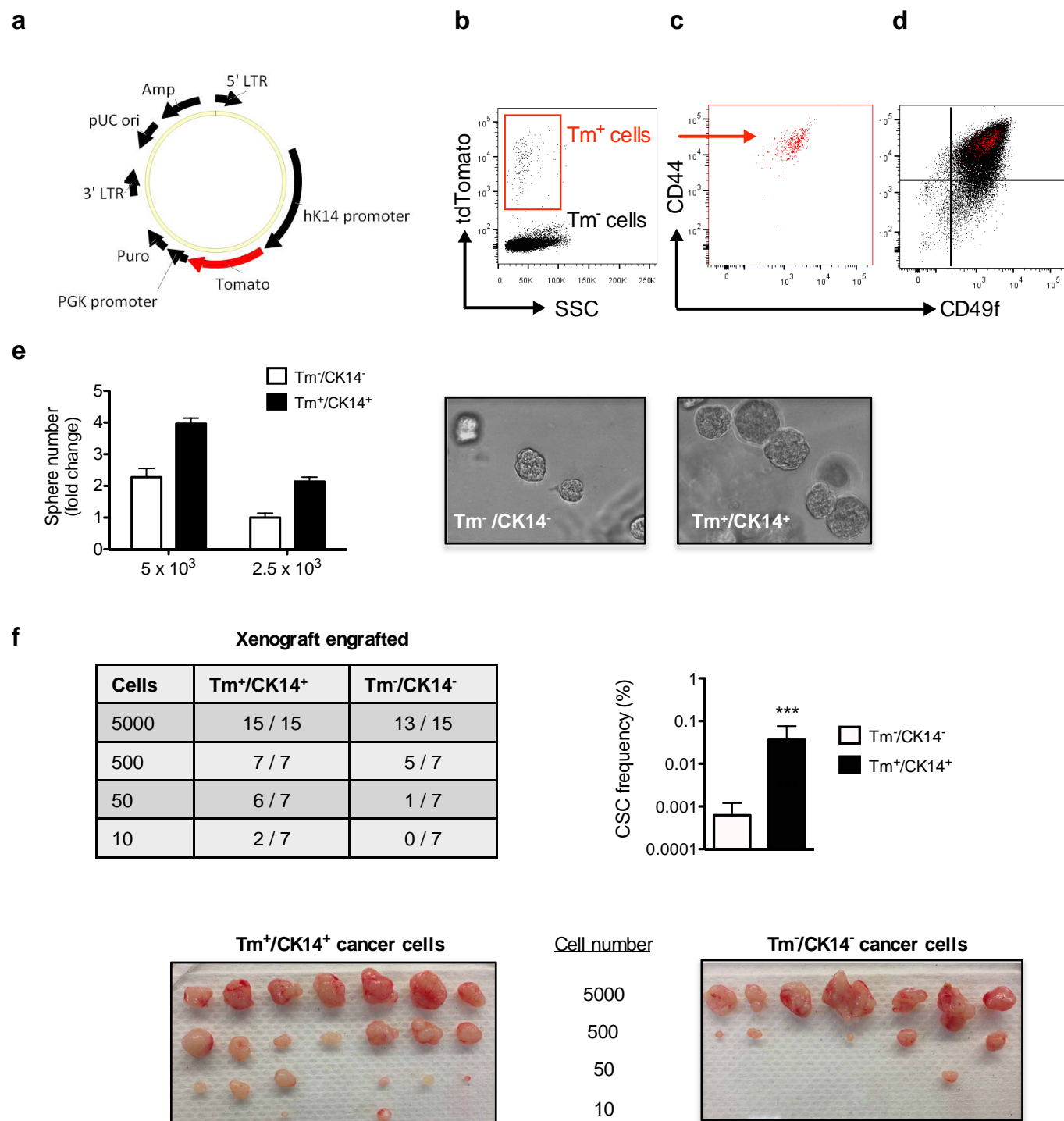
Extended Data Figure 1 | Human bladder urothelial carcinomas and their CK14 status pre- and post-neoadjuvant chemotherapy. **a**, Table summarizing the clinical information of 15 patients with bladder urothelial carcinomas, with paired pre- and post-chemotherapy tissues ($n = 15$). PDX-3 was derived from cancer tissues from patient 3 (bold). **b**, Patient subgroups showing an enrichment (increase) or persistence (maintain high expression) (left) or a resolution (absence) (right) of CK14 staining after chemotherapy

treatment. **c, d**, Representative images of CK14 immunohistochemistry staining (original magnification, $\times 20$), sub-classified as enrichment/persistence (**c**) and resolution (**d**) of CK14 staining in matching bladder urothelial carcinoma tissues obtained pre- and post-neoadjuvant chemotherapy. Representative staining patterns of CK14 high-infiltrative staining (>25%), CK14 low-infiltrative staining (<25%), CK14 basal-restricted staining, and CK14 resolution are individually indicated.



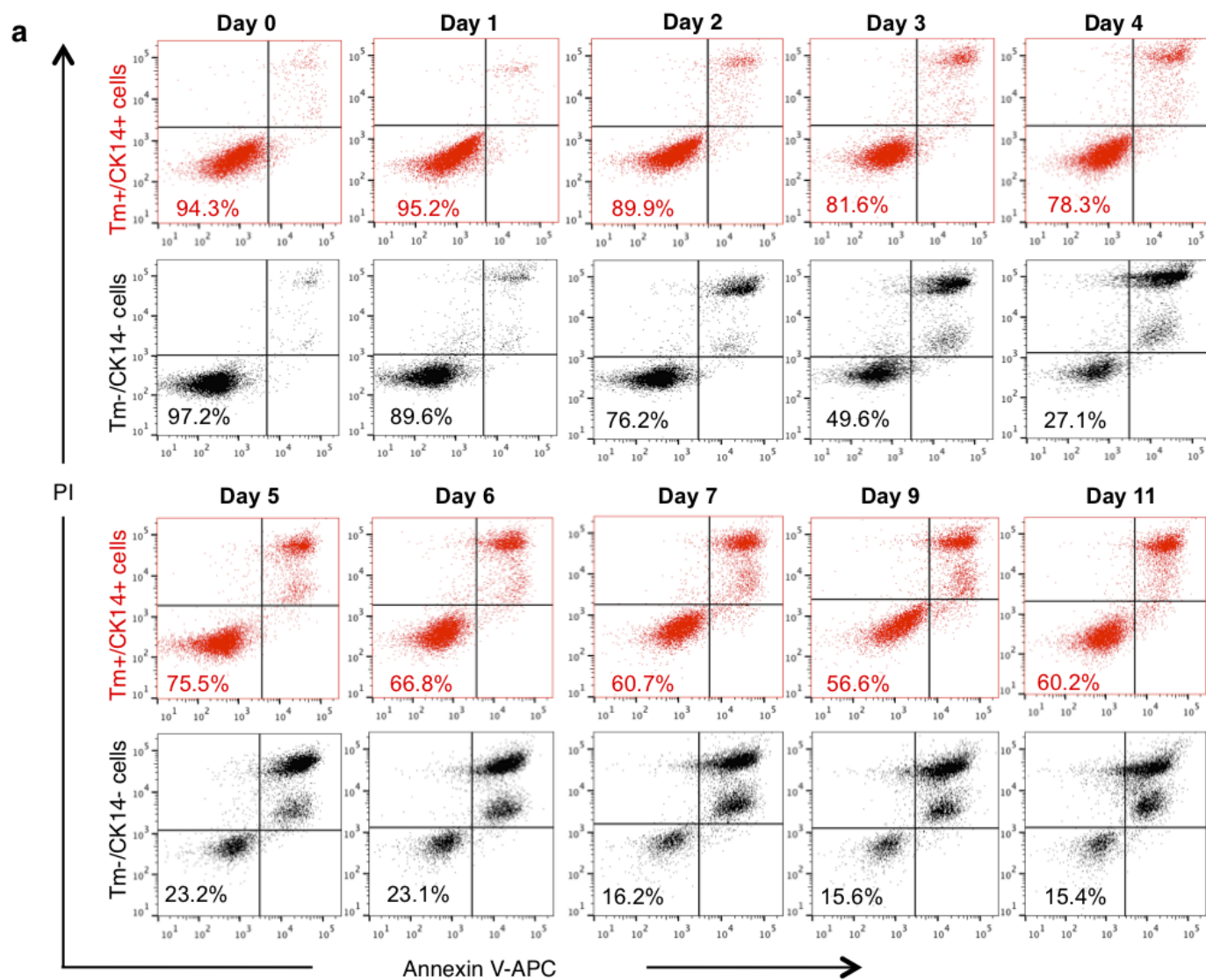
Extended Data Figure 2 | Cytotoxic chemotherapy induces CK14⁺ CSC proliferation despite reducing tumour size. **a**, Relative change in xenograft tumour volume from multiple xenograft tumour lines in response to GC chemotherapy or vehicle treatment ($n = 6$ per group). Xenograft tumours were derived from primary urothelial carcinoma patients (PDX-1 and PDX-2). **b**, Quantification of the percentage of CK14⁺ cancer cells in chemotherapy-treated and vehicle-treated xenograft tumours. **c**, Representative images demonstrating immunofluorescence staining of CK14⁺ cancer cells in chemotherapy-treated and vehicle-treated xenograft tumours.

d, Representative images demonstrating immunofluorescence staining of pHH3 (red) and CK14 (green) in chemotherapy-treated and vehicle-treated xenograft tumours. Yellow arrows indicate CK14⁻pHH3⁺ cells, white arrows indicate CK14⁺pHH3⁺ cancer cells. **e**, Graph quantifying the change in phospho-histone H3 positive (pHH3⁺) proliferating cells within CK14⁺ cancer cells ($n = 6$ per group). All data represent mean \pm s.e.m. Box plots in **b** show twenty-fifth to seventy-fifth percentiles, with line indicating the median and whiskers indicating the smallest and the largest values. * $P < 0.05$; ** $P < 0.01$; *** $P < 0.001$ (two-tailed Student's t -test). Scale bar, 100 μ m.



Extended Data Figure 3 | Methodology to purify CK14⁺ cells for functional evaluation of sphere-forming and tumorigenic properties. **a**, Generation of a reporter construct to isolate viable CK14⁺ bladder cancer cells by FACS. A previously reported and validated human *KRT14* gene promoter fragment was subcloned into a promoterless lentiviral construct that encodes a red fluorescent protein, DD-tdTomato. **b–d**, T24 high-grade urothelial carcinoma cells were stably transduced with the *KRT14* reporter construct. FACS analysis validated that Tm⁺ CK14⁺ cells (red) represent a subpopulation of previously reported CD44⁺ CD49f⁺ tumorigenic cells. **e**, Purified Tm⁺ CK14⁺ cancer

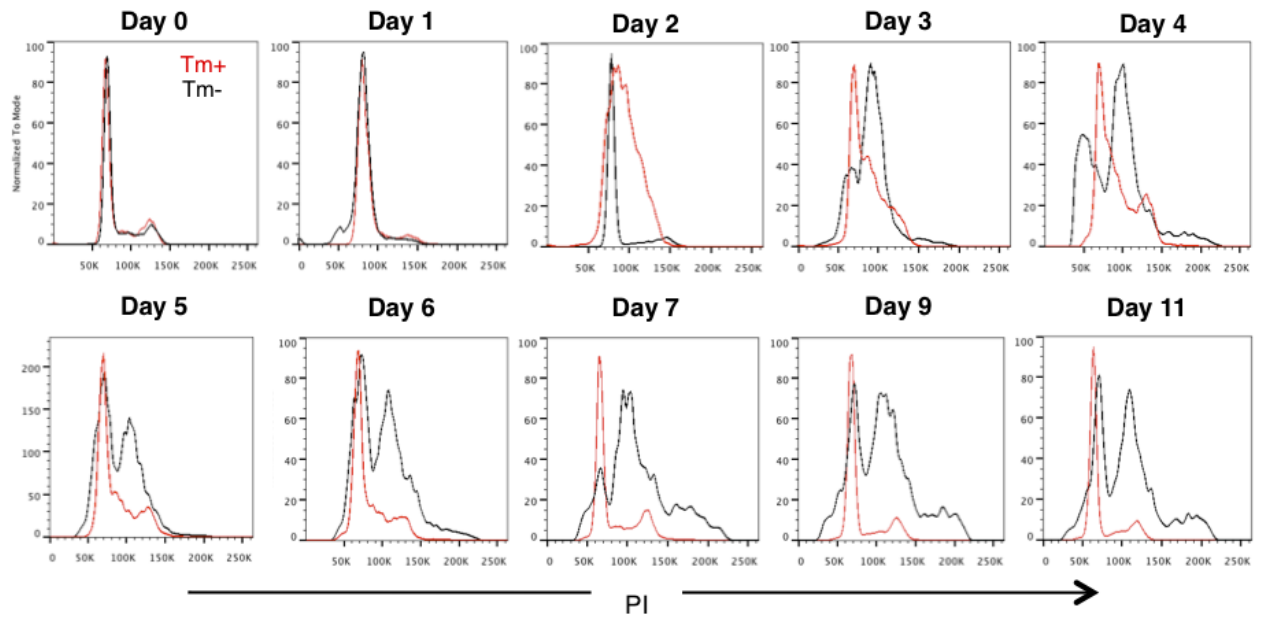
cells demonstrated greater sphere-forming ability than Tm⁻ CK14⁻ cancer cells *in vitro* (biological duplicates). **f**, Tm⁺ CK14⁺ bladder cancer cells are approximately 60-fold enriched for tumorigenic cells when engrafted in immunocompromised mice *in vivo*. Summary of tumour engraftment efficiency and image demonstrating tumour size after transplantation of 10, 50, 500 and 5,000 Tm⁺ CK14⁺ or Tm⁻ CK14⁻ cancer cells as purified by FACS. Data represent mean and range (**e**) and mean \pm s.e.m (**f**). *** $P < 0.001$ (two-tailed Student's *t*-test).



Extended Data Figure 4 | Cell viability of purified Tm⁺ CK14⁺ and Tm⁻ CK14⁻ cancer cells after GC chemotherapy treatment *in vitro* (raw FACS data for Fig. 1h). Dot plots depict FACS analyses showing cell viability of Tm⁺ CK14⁺ (red) and Tm⁻ CK14⁻ (black) cancer cells following

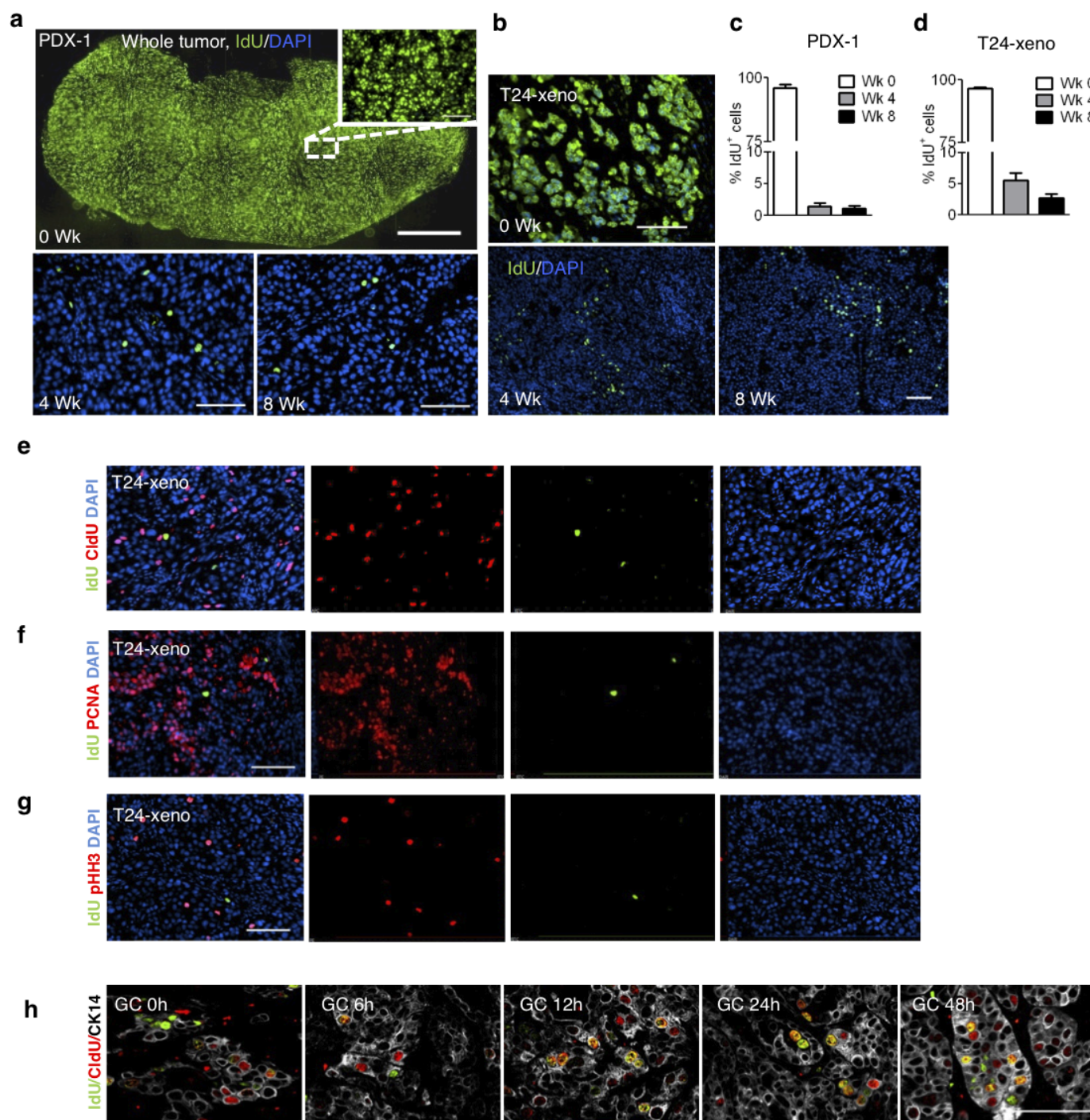
11 consecutive days of chemotherapy treatment *in vitro*. The percentage of viable cells defined as annexin V⁻ PI⁻ is shown in the bottom left quadrant of each plot. Experiments were performed in biological duplicates.

a



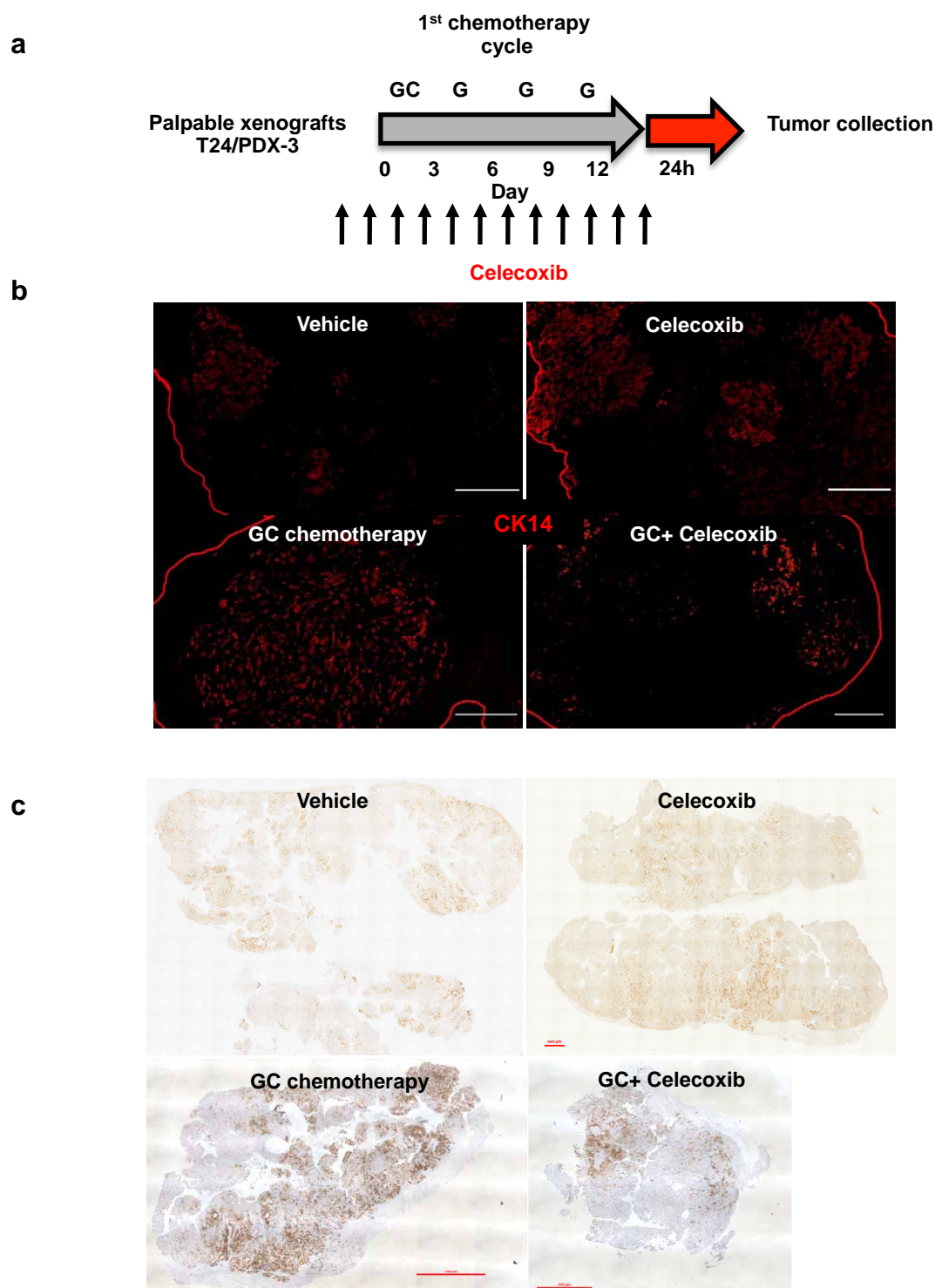
Extended Data Figure 5 | Cell cycle profiles of purified Tm^+ CK14 $^+$ and Tm^- CK14 $^-$ cancer cells after GC chemotherapy *in vitro* (raw FACS data for Fig. 1i). Histogram plots depict original FACS analyses of cell cycle profiles

from Tm^+ CK14 $^+$ (red) and Tm^- CK14 $^-$ (black) cancer cells after 11 consecutive days of chemotherapy treatment *in vitro*. Experiment was performed in biological duplicates. PI, propidium iodide.



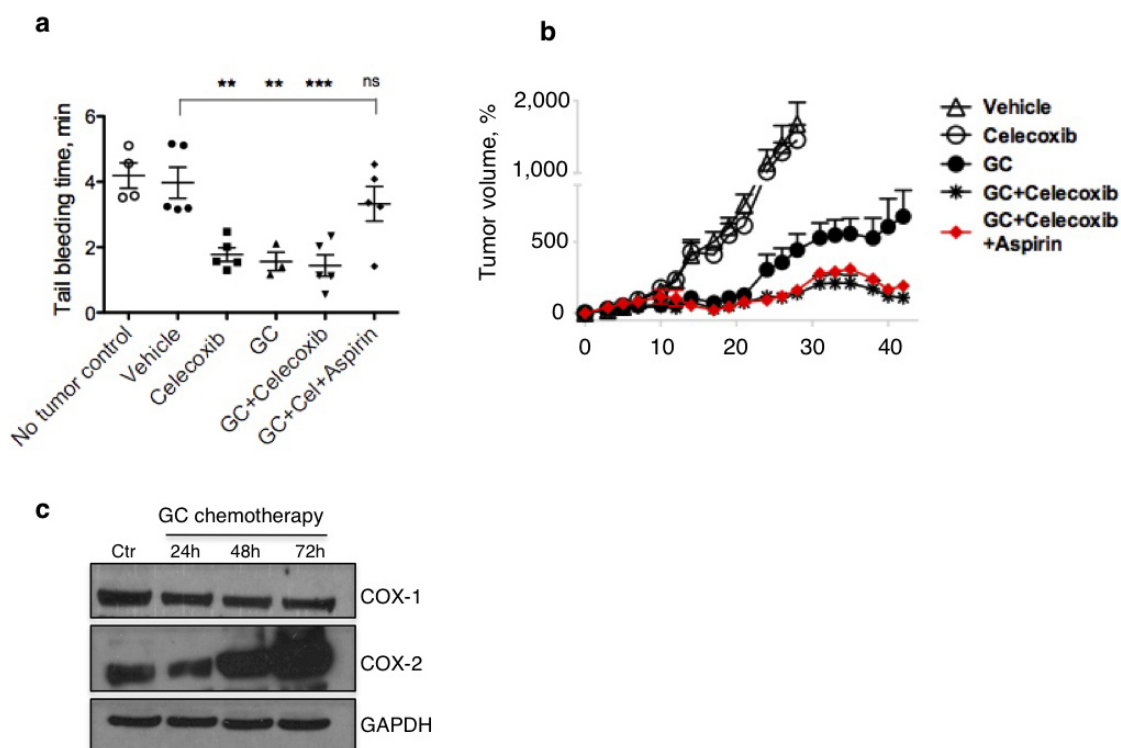
Extended Data Figure 6 | LRCCs are mutually exclusive to active proliferative cancer cells. **a, b**, Immunofluorescence staining to locate LRCCs (green, IdU⁺) at 0, 4 and 8 weeks of chase periods in patient-derived urothelial carcinoma xenograft (PDX-1) (**a**) and xenograft established from T24 high-grade urothelial carcinoma cells (**b**). **c, d**, Bar graph quantifying the percentage of LRCCs in patient-derived xenograft (PDX-1) (**c**) and immortalized cancer xenograft (T24) (**d**) at various chase periods ($n = 4$). **e–g**, Immunofluorescence

staining to evaluate the localization of LRCCs (green, IdU⁺) and proliferating cells (red) using CldU (**e**), proliferating cell nuclear antigen (PCNA; **f**) or phospho-histone H3 (pHH3; **g**) in high-grade urothelial carcinoma (T24) at steady state. **h**, Immunofluorescence co-staining to locate IdU⁺ LRCCs, CldU⁺ proliferating and CK14⁺ cancer cells at various time points after GC chemotherapy. Data shown in **c** and **d** represent mean \pm s.e.m. Scale bars, 100 μ m.



Extended Data Figure 7 | Celecoxib abrogates CK14⁺ cancer cell enrichment after GC chemotherapy in T24 and PDX-3 xenografts. a, *In vivo* treatment protocol recapitulating clinical regimen of one chemotherapy cycle in the presence or absence of celecoxib treatment. **b,** Immunofluorescence staining examining the percentage of CK14⁺ cancer cells in representative T24

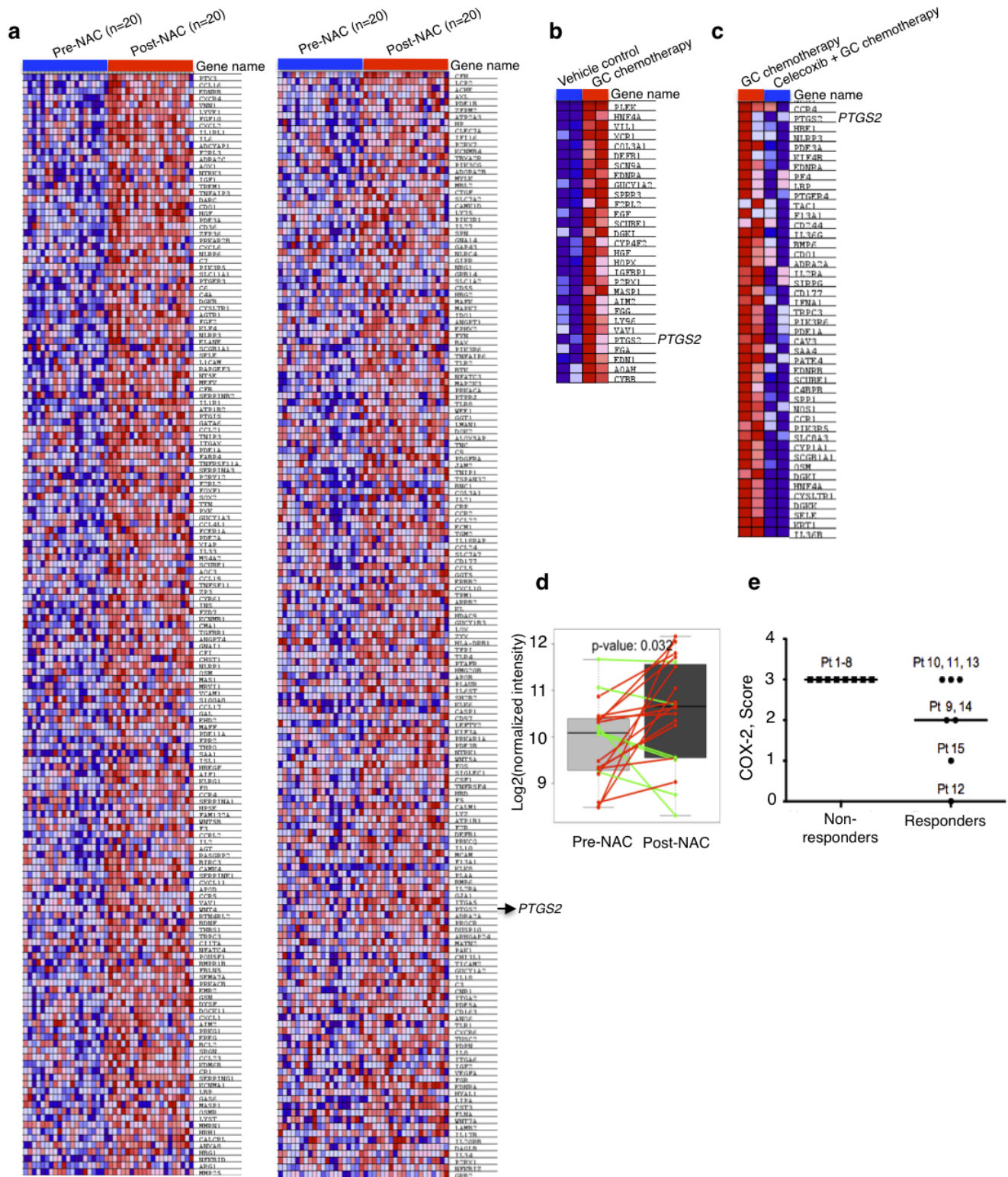
xenograft tumours from various treatment groups. Scale bars, 1,000 μ m. **c,** Immunohistochemical staining examining the percentage of CK14⁺ cancer cells in representative PDX-3 xenograft tumours from various treatment groups. Scale bars, 1,000 μ m. Images in **b** and **c** are representative of $n = 6$ tumours analysed for each treatment group.



Extended Data Figure 8 | Co-administration of aspirin diminishes thrombosis without impairing adjuvant effect of celecoxib.

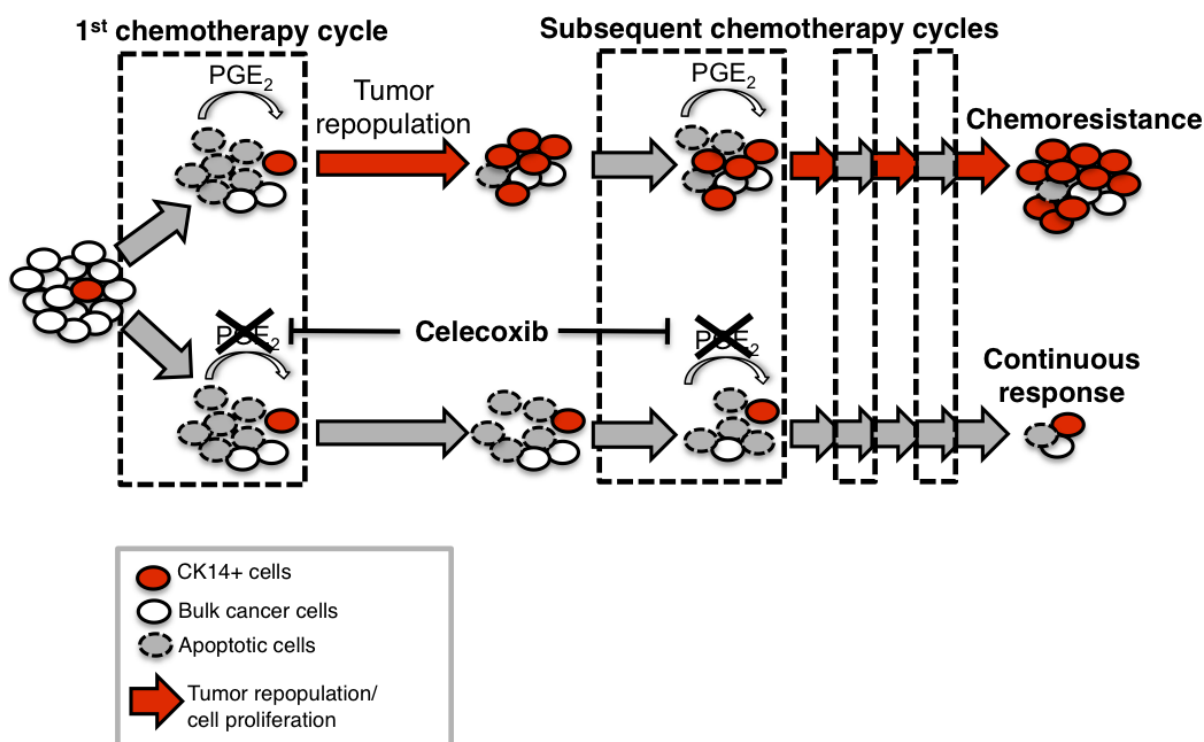
a, Antithrombotic effect of aspirin measurable by tail bleeding time.
b, Temporal percentage change in tumour size after two cycles of celecoxib plus GC combination chemotherapy in the presence or absence of aspirin ($n = 12$ per group). **c**, Dynamics of COX1 and COX2 expression after GC

chemotherapy, data shown for T24 cancer cells (displayed are representative blots from $n = 3$ experiments). All data represent mean \pm s.e.m., tumour volume measurements shown in **b** are relative to mean tumour volume at day 0. * $P < 0.05$; ** $P < 0.01$; *** $P < 0.001$ (one-way ANOVA followed by Dunnett's test for multiple comparisons); ns, not significant.



Extended Data Figure 9 | Enrichment of the 'wound-response gene signature' in chemoresistant bladder urothelial carcinomas. **a**, GSEA validated an enrichment of the 'wound-response gene signature' (GO:0009611) in a panel of non-responding (or chemoresistant) human urothelial bladder carcinomas ($n = 20$; GSE48277), by comparing post-chemotherapy to pre-chemotherapy cancer tissues. Heat map demonstrates part of the genes within leading edge, including the COX2 gene *PTGS2*. **b**, Enlarged heat map for Fig. 4g. **c**, Enlarged heat map for Fig. 4h. **d**, Bioinformatics analysis of chemoresistant

cancers in a panel of non-responding (or chemoresistant) human urothelial bladder carcinomas ($n = 20$; GSE48277) validated a significant increase of *PTGS2* in post-chemotherapy tissues in comparison to matching pre-chemotherapy tissues. **e**, Dot plots representing the scoring of COX2 staining in bladder urothelial carcinoma tissues obtained pre-neoadjuvant chemotherapy in two subgroups of patients with different response to neoadjuvant chemotherapy (subgroups described in Extended Data Fig. 1a–c).



Extended Data Figure 10 | Schematic model: recurrent CSC repopulation and its manifestation of chemoresistance. Cytotoxic chemotherapy effectively induces apoptosis but paradoxically elicits a wound response of bladder cancer stem cells to proliferate and repopulate residual tumours. Release of PGE₂ from neighbouring apoptotic cancer cells is sufficient to

promote this CSC repopulation. *In vivo* administration of celecoxib effectively abolishes this PGE₂/COX2-mediated wound response gene signature, and attenuates progressive manifestation of chemoresistance in preclinical models of human urothelial carcinomas.

T–B-cell entanglement and ICOSL-driven feed-forward regulation of germinal centre reaction

Dan Liu^{1*}, Heping Xu^{1*†}, Changming Shih¹, Zurong Wan¹, Xiaopeng Ma², Weiwei Ma¹, Dan Luo^{1†} & Hai Qi¹

The germinal centre (GC) reaction supports affinity-based B-cell competition and generates high-affinity bone-marrow plasma cells (BMPCs)^{1,2}. How follicular T-helper (T_{FH}) cells regulate GC selection is not clear^{3,4}. Using competitive mixed chimaera, we show here that, beyond the role in promoting T_{FH} development^{5–7}, ICOSL (inducible T-cell co-stimulator ligand, also known as ICOSLG) is important for individual B cells to competitively participate in the GC reaction and to develop into BMPCs. Using intravital imaging aided by a calcium reporter, we further show that ICOSL promotes an ‘entangled’ mode of T_{FH}–B-cell interactions, characterized by brief but extensive surface engagement, productive T-cell calcium spikes, and B-cell acquisition of CD40 signals. Reiterated entanglement promotes outer-zone co-localization of outcompeting GC B cells together with T_{FH} cells, affording the former increased access to T-cell help. ICOSL on GC B cells is up-regulated by CD40 signals. Such an intercellular positive feedback between contact-dependent help and ICOSL-controlled entanglement promotes positive selection and BMPC development, as evidenced by observations that higher-affinity B-cell receptor variants are enriched in the ICOSL^{high} fraction, that numerically disadvantaged ICOSL-deficient GC B cells or BMPCs exhibit strong affinity compensation in competitive chimaera, and that when GC competition proceeds without ICOSL, selection of high-affinity variants in otherwise normal GC reactions is impaired. By demonstrating entanglement as the basic form of GC T_{FH}–B-cell interactions, identifying ICOSL as a molecular linkage between T–B interactional dynamics and positive selection for high-affinity BMPC formation, our study reveals a pathway by which T_{FH} cells control the quality of long-lived humoral immunity.

Immunization studies of *Cd45.2*^{+/+} *Icosl*^{–/–};*Cd45.1*^{+/+} *Icosl*^{+/+} (*Cd45.1* and *Cd45.2* are also known as *Ptprc*^d and *Ptprc*^b, respectively) 50:50 mixed chimaeras clearly reveal a critical role for ICOSL in maintaining the competitiveness of individual B cells to participate in the GC reaction and to develop into BMPCs (Supplementary Note 1 and Extended Data Fig. 1a–j). Using the well-established OT-II/MD4 co-transfer system⁸, we did not observe by intravital imaging any significant difference in contact duration between OT-II cells and wild-type or ICOSL-knockout MD4 cells at the T–B border or in and around GC clusters (Extended Data Figs 1k–l and Supplementary Video 1). However, when T–B contacts in and around GCs were examined under higher magnifications, T cells appeared to interact with knockout and wild-type MD4 B cells differently. As exemplified in Fig. 1a and Supplementary Video 2, T cells more frequently wrapped the wild-type B cells with abruptly extended cellular protrusions, creating a large contact area on which the two cell bodies became briefly ‘entangled’, even without fully stopping migration. Contacts with knockout B cells involved less extensive surface engagement, with the T cells often remaining in a migratory, ‘passing’ shape. To quantitate such differences, on the 2-dimensional z-projection image we calculated the maximal proportion of the T-cell perimeter in contact with the B cell as the ‘surface engagement index’ (SEI) for a given interaction incidence (Fig. 1b). As shown in

Fig. 1c, the average knockout SEI was significantly reduced (mean ± s.e.m., 32 ± 1.0% versus 39 ± 1.0%, *P* < 0.001). This difference did not result from difference in size of the two B-cell populations (Fig. 1d) or irrelevant green fluorescent protein (GFP) or dsRed transgenic effects (Extended Data Fig. 2 and Supplementary Video 3). Knockout and wild-type SEI distributions could be fitted with Gaussian curves intersecting at 38% (Fig. 1e), above which essentially all contacts involved entangled morphologies like the cell highlighted in Fig. 1b. We thus define entangled contacts as those with an SEI above 38% and ‘passing’ contacts as those below. Entangled contacts were short-lived but overall lasted longer than passing contacts by ~1 min (Fig. 1f; mean ± s.e.m., 3.5 ± 0.2 versus 2.5 ± 0.2 min per incidence). T–B entanglement requires T-cell-receptor stimulation because T cells were exclusively in passing mode and exhibited no difference in contact surfaces when encountering individual non-presenting wild-type or knockout B cells (data not shown).

To verify if T–B entanglement is functionally productive, we measured calcium responses by T cells *in vivo* using a fluorescence resonance energy transfer (FRET) calcium reporter, yellow Cameleon (YC)-Nano-50^{CD} (refs 9,10) (see Supplementary Methods, Extended Data Fig. 3 and Supplementary Video 4 for technical details and demonstration). We imaged areas around the apical outer edges of the GC light zone, where GC B cells are not overcrowded and T–B contacts are individually identifiable. The total cytoplasmic YFP fluorescence tracked migrating T cells; the fractional change in the normalized FRET signal intensity, $\Delta R_t/R_0$, reflects the change in intracellular Ca²⁺ concentration at time *t* as compared to the no-contact reference state 0 of the same T cell before a contact. As exemplified in Fig. 1g and Supplementary Video 5a, b, significant calcium fluxes were observed in T cells during entangled but not passing contacts. Using a stringent threshold of the mean $\Delta R_t/R_0$ being no less than 50%, pronounced calcium spikes were observed in 28 out of 49 (57%) entangled contacts, including those lasting for no more than 2 min, and in 2 out of 39 (5%) contacts in the passing mode (Fig. 1i). Thus, T–B entanglement is associated with productive triggering of the T cells. Next, we imaged the calcium response of T cells contacting *Icosl*^{–/–} GC B cells. As exemplified in Fig. 1h and Supplementary Video 5c, entangled contacts with knockout B cells were associated with detectable but subdued calcium fluxes in the T cells (Fig. 1j; mean ± s.e.m. of $\Delta R_t/R_0$ for 49 wild-type versus 30 knockout contacts, 52 ± 6% versus 32 ± 7%; *P* < 0.05). To gauge the overall productivity of contact events involving the two B-cell types, the mean $\Delta R_t/R_0$ of each contact was multiplied by its duration to obtain the area under the curve as the time-integrated calcium response index (TICRI; Fig. 1k). As shown in Fig. 1l, while TICRIs of passing contacts were comparably very low regardless of the ICOSL expression (knockout, 31 ± 10; wild type, 32 ± 12; *P* = 0.9), TICRIs of entangled contacts with wild-type B cells were close to three times as high as those with knockout B cells (wild type, 326 ± 47; knockout, 124 ± 26; *P* < 0.001). Therefore, short but entangled T–B contacts are functional, and their productivities are markedly reduced without ICOSL, a conclusion further supported by additional imaging analyses of the T-cell

¹Tsinghua-Peking Center for Life Sciences, Laboratory of Dynamic Immunobiology, School of Medicine, Tsinghua University, 100084 Beijing, China. ²School of Life Sciences, Tsinghua University, 100084 Beijing, China. [†]Present addresses: Division of Immunobiology and the Center for Systems Immunology, Cincinnati Children’s Hospital Medical Center, Cincinnati, Ohio 45229, USA (H.X.); College of Life Sciences & Bioengineering, Jiaotong University, 100044 Beijing, China (D. Luo).

*These authors contributed equally to this work.

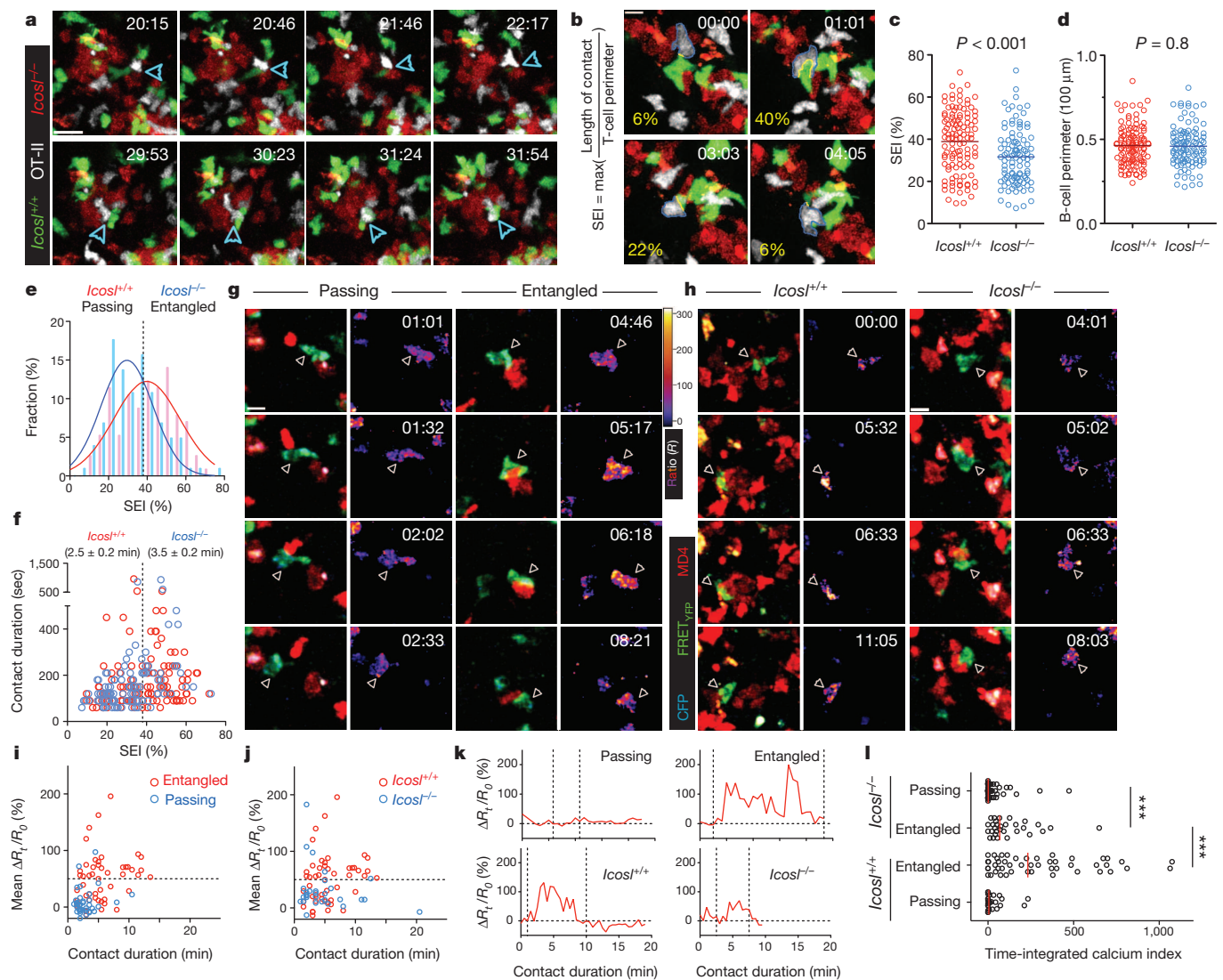


Figure 1 | GCT-B entanglement controlled by ICOSL and ICOS. **a–f**, OT-II T cells, *Icosl*^{+/+} and *Icosl*^{-/-} MD4 B cells visualized 96–120 h after hen egg lysozyme (HEL)-OVA immunization. **a**, Typical ‘entangled’ and ‘passing’ behaviours (arrowheads). See corresponding Supplementary Video 2. Scale bar, 20 μm. **b**, Surface engagement index (SEI). Numbers in yellow denote the proportion of the T-cell perimeter (blue) engaged in contact with the B cell (yellow). The maximal proportion reached during a contact is defined as its SEI. Scale bar, 10 μm. **c**, SEI scatter-plots for 114 *Icosl*^{+/+} and 102 *Icosl*^{-/-} B cells. **d**, Perimeters of analysed B cells. Lines denote the means. **e**, **f**, SEI histograms and fitted Gaussian curves (R^2 for goodness of fit: *Icosl*^{-/-}, 0.85; *Icosl*^{+/+}, 0.80) (**e**) and plot of contact duration against SEI (**f**); the dotted line marks 38%, at which the two curves intersected. Numbers in parentheses are mean \pm s.e.m. contact duration. **g–l**, Calcium imaging of YC-Nano-50^{CD}-transduced OT-II T cells contacting *Icosl*^{+/+} or *Icosl*^{-/-} dsRed MD4 B cells

96–120 h post-immunization. **g**, **h**, Fluorescence overlay (left) or pseudo-colour FRET ratio (R) images (right) of OT-II cells in passing or entangled contact with MD4 cells (*Icosl*^{+/+} or *Icosl*^{-/-} MD4 cells (**h**). Scale bar, 10 μm; CFP, cyan fluorescent protein; YFP, yellow fluorescent protein. **i**, Plot of the mean $\Delta R_t/R_0$ against contact duration of 39 passing and 49 entangled contacts. Data pooled from three experiments. The dotted line indicates mean $\Delta R_t/R_0 = 50\%$. **j**, Plot of the mean $\Delta R_t/R_0$ against contact duration of entangled contacts with 49 *Icosl*^{+/+} or 30 *Icosl*^{-/-} MD4 B cells. Data for *Icosl*^{-/-} cells pooled from two experiments. The dotted line indicates mean $\Delta R_t/R_0 = 50\%$. **k**, $\Delta R_t/R_0$ traces over time for the highlighted OT-II cells in **g** and **h**; dotted vertical lines indicate the contact duration. **l**, Time-integrated calcium indices (TICRI) of all genotypes and contact types. Red lines denote means; *** $P < 0.001$. See corresponding Supplementary Video 5.

calcium response in the densely packed light zone core (Supplementary Note 2, Extended Data Fig. 4 and Supplementary Video 6).

To test whether ICOSL co-signalling promotes the delivery of the contact-dependent help signal CD40L (CD40 ligand), we examined CD40L mobilization from the intracellular store of effector T cells^{11,12}. As shown in Extended Data Fig. 5, within 3–5 min of *ex vivo* T-cell receptor stimulation, a time scale consistent with the brevity of T–B entanglement in GCs, ICOS (inducible T-cell co-stimulator) co-signalling enhanced CD40L externalization to the cell surface. Moreover, its effect can be replicated by a calcium ionophore. Therefore, ICOSL–ICOS interactions not only promote the extent of surface engagement during brief, entangled T–B contacts but also promote the rapid surface display of CD40L molecules.

Consistent with these effects, messenger RNA sequencing analyses of ICOSL-knockout and wild-type GC B cells sorted from the same animals indicated that knockout cells were significantly deprived of transcriptional imprints of CD40 signalling (Extended Data Fig. 6). We conclude that T-cell help in GCs is mainly delivered through entangled contacts, those that are brief, non-stable, but involve extensive cell surface engagement and rapid CD40L mobilization specifically controlled by ICOSL–ICOS co-stimulation and calcium signalling.

Access to T_{FH} cells is limiting in GCs^{3,4}. To test whether differential entanglement as a result of different ICOSL expression would influence how B cells compete for GC T_{FH} cells, we imaged dynamic behaviours of *Icosl*^{+/+} MD4 GC B cells facing a numerical majority of competing

Icosl^{+/+} or *Icosl*^{-/-} MD4 GC B cells, which differ in the ability to entangle T_{FH} cells but have identical, fixed B-cell receptor (BCR) specificity. The tester *Icosl*^{+/+} MD4 B cells and the competing cells were transferred at a 1:40 ratio. As a result, the entire GC proper could be visualized as a 'solid' fluorescent mass of the competing population, whereas individual tester cells were readily identified. Our imaging fields were chosen so as to contain large numbers of T cells inside the GC, largely corresponding to the light zone. As shown in Supplementary Video 7, around 96 h post-immunization when GCs initially formed, the tester GC B cells were scattered throughout the GC proper, while the OT-II T_{FH} cells were also evenly distributed. However, as the competition with the underperforming knockout GC cells progressed for approximately 24 h, the tester MD4 B cells began to be localized to the outer area of the light zone (Fig. 2a and Supplementary Video 8). By contrast, the tester cells remained evenly distributed throughout the light zone when competing with wild-type GC B cells that can equally entangle with T_{FH} cells (Fig. 2a and Supplementary Video 9). Quantitative analyses of the edge localization index (ELI), which is the normalized distance to the closest

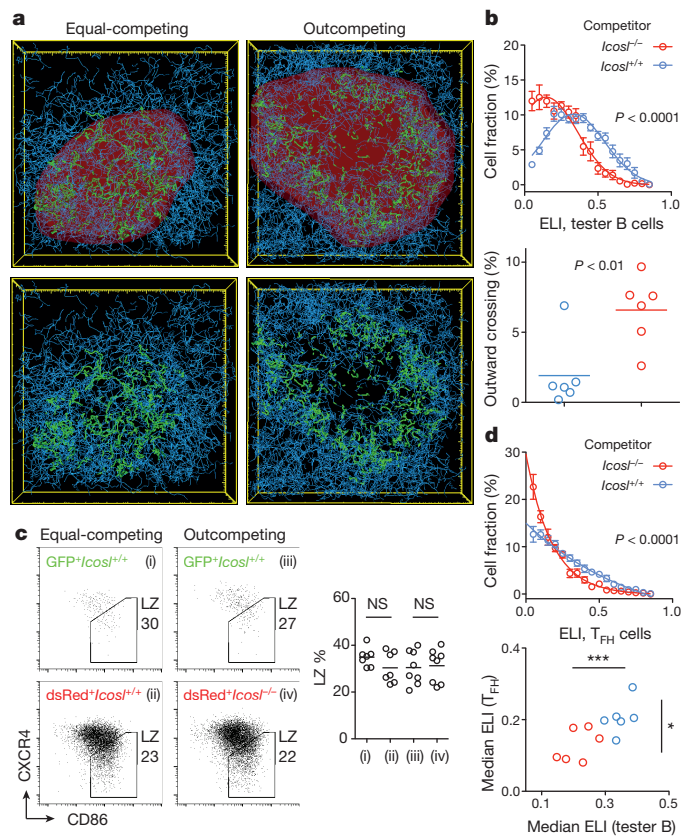


Figure 2 | Coordinated outer-zone localization of outcompeting *Icosl*^{+/+} B cells and associated T_{FH} cells. **a**, *Icosl*^{+/+} tester MD4 B cell tracks (green) and OT-II T_{FH} tracks (blue) in GCs composed of a majority of *Icosl*^{+/+} (left) or *Icosl*^{-/-} (right) competing MD4 cells. The GC proper is highlighted as a red isosurface (top) or left unmarked (bottom). Patterns represent six mice imaged in three experiments. See corresponding Supplementary Videos 8 and 9. **b**, Top panel, ELI histograms of tester B cells and fitted Gaussian curves (R^2 for goodness of fit: *Icosl*^{-/-}, 0.84; *Icosl*^{+/+}, 0.79; P calculated by extra sum of squares F -test; see Extended Data Fig. 7 for ELI quantitation). Bottom panel, outward GC edge-crossing fractions of tester B cells. Lines denote means. **c**, Frequencies of the light zone (LZ) phenotype in tester and competitor B cells under the two conditions. Data pooled from two experiments; NS, not significant. **d**, Top panel, ELI histograms of T_{FH} cells and exponential decay-fitted curves (R^2 for goodness of fit: *Icosl*^{-/-}, 0.89; *Icosl*^{+/+}, 0.87; P calculated by extra sum of squares F -test). Bottom panel, population ELI medians of tester B and corresponding T_{FH} cells. Error bars in **b** and **d** are s.e.m. of 6 or 5 independent experiments. * $P < 0.05$, *** $P < 0.001$.

GC edge (see Extended Data Fig. 7 for details), and the outward border-crossing index, which is the fraction of cell tracks that crossed the GC border from inside to out, demonstrated that outcompeting tester B cells in the light zone were concentrated towards and more often outwardly crossed the GC edge (Fig. 2b). These data suggest that a GC-exiting posture is induced in some GC B cells following competitive acquisition of sufficient T-cell help. At the same time, the overall light-zone–dark-zone distribution of the tester cells did not significantly change between the two conditions of competition (Fig. 2c), consistent with continued GC recycling⁴. Strikingly, as the tester GC B cells became distributed towards the outer zone under the outcompeting but not the equal-competing condition, so did many OT-II T_{FH} cells (Fig. 2a, d, e and Supplementary Videos 8 and 9). The coordinated outer-zone distribution of T_{FH} cells and GC B cells that have outcompeted for T-cell help is reminiscent of human tonsil GCs, in which outer zones are enriched in plasmacytoid cells and T cells with abundant intracellular CD40L¹¹. Therefore, although ICOSL-controlled T–B entanglement is brief, when reiterated, it could facilitate a small number of GC B cells with an advantage in entangling T_{FH} cells to dominate access to T-cell help.

These results raise the interesting possibility that, rather than simply conferring a basal competence to entangle with T_{FH} cells, ICOSL on individual GC B cells may be dynamically regulated in positive correlation with their acquisition of T_{FH} help signals (for example, CD40L) and/or BCR signals, potentially giving rise to feed-forward dynamics that can focus help signals to high-affinity B-cell variants and to drive BMPC development. To test this, we compared ICOSL expression on naive and GC B cells and found that individual GC B cells expressed lower but much more varied levels of ICOSL than naive B cells (Fig. 3a). When freshly isolated B cells containing both GC and non-GC B cells were stimulated through the BCR or CD40 receptor, GC but not non-GC B cells uniquely responded to CD40 stimulation by ICOSL upregulation (Fig. 3b). BCR stimulation actually downregulated ICOSL on both GC and non-GC B cells. CD40-mediated ICOSL upregulation on GC B cells was also observed *in vivo* (Fig. 3c). Furthermore, acutely increased provision of a full complement of T-cell help signals, as achieved by infusion of OVA₃₂₃ peptide to mice immunized with 4-hydroxy-3-nitrophenylacetyl hapten (NP)-OVA in the presence of transferred OT-II cells, also increased ICOSL expression on GC B cells *in vivo* (Fig. 3d). Therefore, T-cell help in the form of CD40L can drive ICOSL upregulation on GC B cells *in vivo*.

It seems that an intercellular feed-forward loop between ICOSL-controlled acquisition of contact-dependent T-cell help and the help-promoted ICOSL upregulation may regulate GC competition and promote positive selection. Such a model predicts that increases in the BCR affinity could be positively translated into increased ICOSL levels through reiterative entanglement with T cells, and thus high-affinity mutants should more likely be found in GC B cells that express the highest levels of ICOSL. We thus analysed the primary NP-specific response in B6 mice, in which V_H186.2-carrying B-cell clones dominate the GC, while a W33L mutation in the V_H confers a tenfold increase in the NP-specific affinity¹³ and is frequently found in the BMPC compartment¹⁴. NP-binding GC B cells expressing the highest (top 10%) and the lowest (bottom 25%) ICOSL levels were sorted 9.5 days post-immunization and subjected to V_H186.2 sequence analyses. As shown in Fig. 3e, 42 out of 94 unique sequences harboured the W33L mutation in the ICOSL^{high} group, while only 14 out of 67 ICOSL^{low} clones were found to carry this mutation ($P = 0.001$), whereas the total mutation loads were comparable between the two (mean \pm s.e.m., 3.5 ± 0.2 versus 3.2 ± 0.2 , $P = 0.3$).

The ICOSL-driven feed-forward model also suggests that high-affinity BCRs should compensate for the disadvantage that ICOSL-knockout GC B cells experience while competing against wild-type counterparts for T-cell help in the same host. To test this, we analysed BCR sequences of NP-binding GC B cells from mixed chimaeras. As shown in Extended Data Table 1, to gauge the counter-selection against ICOSL-knockout cells, we divided the ratio of W33L frequencies in knockout and wild-type cells of the knockout:wild-type chimaera by the same ratio in the parallel wild-type:wild-type chimaera and obtained a counter-selection

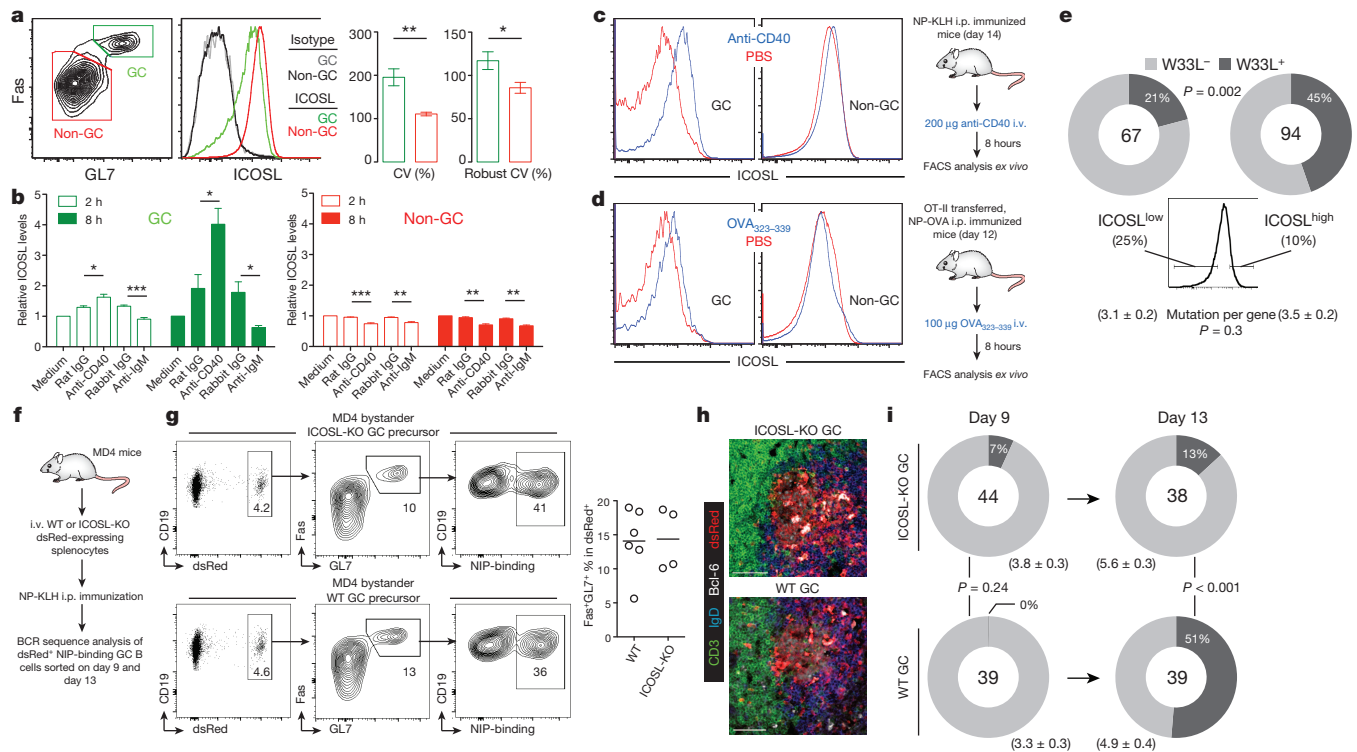


Figure 3 | Dynamic ICOSL regulation and ICOSL-driven feed-forward GC selection. **a**, Staining profiles, coefficient of variance (CV) and robust CV (50 × (intensity at 84.13 percentile – intensity at 15.87 percentile) ÷ median) of ICOSL expression on splenic GC and non-GC B cells. Data are mean ± s.e.m. of three mice (**P* < 0.05; ***P* < 0.01) from one of more than three similar experiments. **b**, Fold changes of surface ICOSL expression on GC or non-GC B cells that were stimulated with anti-IgM or anti-CD40 or isotype-matched control antibodies for 2 or 8 h, with medium control set as unit 1. Data presented are mean ± s.e.m. of 4 (2 h) or 3 (8 h) independent experiments (**P* < 0.05; ***P* < 0.01; ****P* < 0.001). **c, d**, *Ex vivo* ICOSL expression on splenic GC or non-GC B cells 8 h after injection with 200 µg anti-CD40 antibody or PBS into NP-KLH-immunized B6 mice (c) or after injection with 100 µg OVA_{323–339} peptide into OT-II-transferred, NP-OVA-immunized B6 mice (d). Data represent three similar experiments. i.v., intravenous; i.p.,

intraperitoneal. **e**, W33L frequencies among unique V_H186.2 sequences (total numbers at the centre of the circles) of ICOSL^{high} (top 10%) and ICOSL^{low} (bottom 25%) NP-specific GC B cells in B6 mice 9.5 days after NP-KLH immunization. Data are pooled from 3 experiments (*P* calculated by Fisher's exact test; mutation rates in mean ± s.e.m.). **f–i**, GC competition in the presence or absence of ICOSL expression by GC B cells. **f**, The protocol to induce NP-specific GCs from ICOSL-knockout (KO) or wild-type (WT) precursors. **g**, Gates for sorting dsRed-expressing, NP-specific GC B cells (left) and the magnitude of GC responses by ICOSL-knockout or wild-type precursors at day 13 (right); lines denote means. **h**, Representative histology of the two types of GCs. Scale bars, 50 µm. **i**, Mutation rates (mean ± s.e.m.) and W33L frequencies among V_H186.2 sequences of NP-specific GC B cells at day 9 and day 13 (*P* calculated by Fisher's exact test). One of two sets of experiments with similar results is shown.

index of 1.9, which indicates strong selection against ICOSL-knockout cells in the mixed GC. When the BMPC compartment was analysed (Extended Data Table 2), the counter-selection index was 2.5, revealing even stronger counter-selection against and affinity-compensation for the ICOSL deficiency. Notably, in both the GC and the BMPC compartments of the knockout:wild-type chimaera, wild-type cells carried a lower mutation load, suggesting that these advantaged cells experience not only relaxed selection but probably also a shortened GC residence while being positively selected for output (Supplementary Notes 3 and 4). Therefore, in both the GC and BMPC compartments of knockout:wild-type chimaeras, numerically disadvantaged knockout cells are indeed of higher comparative affinities because of compensation for counter-selection.

To test whether the ICOSL-driven feed-forward mechanism is required for positive selection and affinity maturation in the GC, we transferred polyclonal dsRed-expressing wild-type or ICOSL-knockout B6 splenocytes into MD4 recipient mice that were subsequently immunized with NP-KLH (keyhole limpet haemocyanin) (Fig. 3f), thereby creating a system in which the majority of follicular bystander B cells express ICOSL but antigen-specific GC precursor B cells do or do not. Consistent with the fact that ICOSL–ICOS interactions between antigen-specific T and B cells do not appreciably impinge on long-lasting, stable T–B conjugation at the T–B border (Extended Data Fig. 1k) and are not strictly required for the formation of GCs¹⁵, GCs in both settings developed (Fig. 3g), comparably harboured GC T_{FH} cells (Fig. 3h), and accumulated

mutations over time (Fig. 3i). Notably, however, the selection for otherwise dominant high-affinity W33L mutants became significantly slower and inefficient during GC competition in the absence of ICOSL (Fig. 3i). Therefore, the ICOSL-driven feed-forward loop is required for optimal positive selection and affinity maturation in the GC (Supplementary Note 5; see Extended Data Fig. 8 for a schematic representation).

The brevity of T–B entanglement, probably dictated by intrinsic T_{FH} properties and GC microenvironment, particularly suits the need to discriminate B cells in competition, because T cells may gain a larger dynamic range of help signal output by using the surface contact area and ICOSL co-signalling intensity as variables during reiterative interactions with B cells. The identification of a receptor–ligand pair on the cell surface, ICOSL and ICOS, that specifically promotes GC selection in a feed-forward fashion may lead to novel strategies for tailoring affinity features of vaccine-induced responses in the future.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 29 October 2013; accepted 27 August 2014.

Published online 15 October 2014.

1. Victora, G. D. & Nussenzweig, M. C. Germinal centers. *Annu. Rev. Immunol.* **30**, 429–457 (2012).
2. Tarlinton, D. & Good-Jacobson, K. Diversity among memory B cells: origin, consequences, and utility. *Science* **341**, 1205–1211 (2013).

3. Allen, C. D., Okada, T., Tang, H. L. & Cyster, J. G. Imaging of germinal center selection events during affinity maturation. *Science* **315**, 528–531 (2007).
4. Victora, G. D. *et al.* Germinal center dynamics revealed by multiphoton microscopy with a photoactivatable fluorescent reporter. *Cell* **143**, 592–605 (2010).
5. Gigoux, M. *et al.* Inducible costimulator promotes helper T-cell differentiation through phosphoinositide 3-kinase. *Proc. Natl Acad. Sci. USA* **106**, 20371–20376 (2009).
6. Choi, Y. S. *et al.* ICOS receptor instructs T follicular helper cell versus effector cell differentiation via induction of the transcriptional repressor Bcl6. *Immunity* **34**, 932–946 (2011).
7. Xu, H. *et al.* Follicular T-helper cell recruitment governed by bystander B cells and ICOS-driven motility. *Nature* **496**, 523–527 (2013).
8. Qi, H., Cannons, J. L., Klauschen, F., Schwartzberg, P. L. & Germain, R. N. SAP-controlled T–B cell interactions underlie germinal centre formation. *Nature* **455**, 764–769 (2008).
9. Horikawa, K. *et al.* Spontaneous network activity visualized by ultrasensitive Ca^{2+} indicators, yellow Cameleon-Nano. *Nature Methods* **7**, 729–732 (2010).
10. Mues, M. *et al.* Real-time *in vivo* analysis of T cell activation in the central nervous system using a genetically encoded calcium indicator. *Nature Med.* **19**, 778–783 (2013).
11. Casamayor-Palleja, M., Khan, M. & MacLennan, I. C. A subset of CD4^{+} memory T cells contains preformed CD40 ligand that is rapidly but transiently expressed on their surface after activation through the T cell receptor complex. *J. Exp. Med.* **181**, 1293–1301 (1995).
12. Koguchi, Y., Thauland, T. J., Slifka, M. K. & Parker, D. C. Preformed CD40 ligand exists in secretory lysosomes in effector and memory CD4^{+} T cells and is quickly expressed on the cell surface in an antigen-specific manner. *Blood* **110**, 2520–2527 (2007).
13. Allen, D., Simon, T., Sablitzky, F., Rajewsky, K. & Cumano, A. Antibody engineering for the analysis of affinity maturation of an anti-hapten response. *EMBO J.* **7**, 1995–2001 (1988).
14. Smith, K. G., Light, A., Nossal, G. J. & Tarlinton, D. M. The extent of affinity maturation differs between the memory and antibody-forming cell compartments in the primary immune response. *EMBO J.* **16**, 2996–3006 (1997).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank L. and M. McHeyzer-Williams for providing the V_H gene analysis protocol, T. Nagai for YC-Nano reporter constructs⁹, and H. Wekerle for the Twitch-1^{CD} reporter construct¹⁰. H.Q. is indebted to Y. Hong for support. H.Q. was a Tsinghua–Bayer Investigator and a Tsinghua–Janssen Investigator. This work was funded in part by the Ministry of Science and Technology ‘973’ program (2014CB542501), National Natural Science Foundation of China (81330070 and 81361120397), Tsinghua University Initiative Scientific Research Program (20131089224), Institut Mérieux, and the Tsinghua–Peking Center for Life Sciences.

Author Contributions D. Liu and H.X. conducted a majority of the experiments; C.S. developed the intravital calcium imaging method and conducted much of the imaging work together with H.X.; Z.W. developed the assay for ICOSL-deficient GCs and conducted related mutational analyses; X.M. performed RNA-seq analysis; W.M. and D. Luo helped with imaging and molecular cloning. H.Q. conceptualized the study, designed the experiments with input from H.X., D. Liu, C.S. and Z.W., and wrote the paper. All authors contributed collectively to interpreting data.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to H.Q. (qihai@tsinghua.edu.cn).

METHODS

Mice. B6 (Jax 664), CD45.1 (Jax 002014), GFP-expressing (Jax 4353), CFP-expressing (Jax 4218), dsRed-expressing (Jax 6051), *Icosl*^{−/−} (Jax 4657), OVA_{323–339}-specific T-cell receptor transgenic OT-II (Jax 4194), and HEL-specific Ig-transgenic MD4 (Jax 2595) mice were from the Jackson Laboratory. Relevant mice were interbred to obtain CFP OT-2, *Icosl*^{+/+} GFP MD4, and *Icosl*^{+/+} or *Icosl*^{−/−} dsRed MD4 mice. All mice were maintained under specific-pathogen free conditions, and used in accordance with governmental and institutional guidelines for animal welfare. Age- and sex-matched donor or recipient mice were randomly chosen to be included in different treatment groups for comparison, typically involving 4 or 5 per group, and completely independent animal studies were always repeated at least twice, but up to 6 times. No blinding was necessary for animal experiments involved in this study.

Immunization. To measure NP-specific response in B6 mice or various chimaeras, the animals were intraperitoneally injected with 100 µg NP-KLH (Biosearch Technologies) mixed with 1 µg lipopolysaccharide (LPS) (Sigma) in alum (Thermo Scientific). To visualize cognate T–B interactions at the T–B border and in the GC or to conduct RNA sequencing (RNA-seq) analysis of MD4 GC B cells, recipients of OT-II T cells and MD4 B cells of indicated genotypes were subcutaneously immunized with a suspension of 0.5 µg LPS and alum in PBS supplemented with 30 µg HEL-OVA, which was made by chemical crosslinking with a HydraLink conjugation kit (SoluLink) as previously described¹⁶.

Cell culture, retrovirus and *in vitro* transduction. Naive T cells or B cells were isolated using the Negative CD4 T-cell Isolation Kit or the Naive B-cell Isolation Kit (Miltenyi Biotec) according to the manufacturer's protocols. Retroviral transduction of mouse lymphocytes was done as previous described⁷.

Flow cytometry. Spleen, lymph node or bone marrow cells were washed, blocked with a mixture of 10% goat and rabbit serum and 20 µg ml^{−1} 2.4G2 antibody (BioXcell), and then stained with antibodies of indicated specificities in MACS buffer. Staining reagents included AlexaFluor700 anti-CD4, APC-Cy7 anti-CD19, phycoerythrin (PE) anti-ICOSL, streptavidin–allophycocyanin (APC), streptavidin–PE and biotinylated anti-ICOSL purchased from Biolegend; eFlour450 anti-IgD, eFlour450 anti-IgM, eFlour450 anti-CD45.2, PE-Cy7 anti-CD45.1, FITC GL7, AlexaFluor700 anti-CD38 and AlexaFluor700 anti-B220 from eBioscience; PE anti-CD95, PE anti-CD40L, PE-Cy7 anti-CD95, biotinylated anti-IgM⁺, APC anti-CD138 and V450 anti-Gr1 from BD Biosciences; NP(8)-conjugated PE (Biosearch Technologies) was used to stain for NP-binding cells. Isotype-matched non-specific antibodies were from the same sources. Cells were typically stained on ice with primary reagents for 60–90 min followed by staining with secondary reagents for 30 min. Data were collected on an LSR II cytometer (BD) and analysed with FlowJo software (TreeStar). Dead cells and cell aggregates were excluded from analyses based on 7-AAD staining (Biotium) and FSC-H/FSC-A characteristics.

Mixed bone-marrow chimaeras and measurement of competitive competencies. CD45.2 B6 mice were lethally irradiated by X-ray (500 rad × 2) and then intravenously transferred with 4 × 10⁶ bone-marrow leukocytes from donors of indicated genotypes at 1:1 ratio. Chimeras were used for immunization 8 weeks after reconstitution. Without enriching for stem cells, the final ratio between the genotypes in lymphocytes of the chimaeric animals can significantly deviate from 1:1. To measure the competency of *Icosl*^{−/−} or *Icosl*^{+/+} CD45.2 B cells in competition with CD45.1 B cells to become GC B cells (CD19⁺Fas⁺GL7⁺), BMPCs (BM B220^{low} CD138^{high}), or antigen-specific memory B cells (NP-binding⁺ IgD[−] IgM[−] CD4[−] CD19⁺ CD38⁺), the ratio of CD45.2 and CD45.1 cells in each compartment for each immunized mouse was individually normalized against the CD45.2:CD45.1 ratio in the splenic reference B-cell compartment of the same mouse. For convenience of combined staining and efficient utilization of splenic tissues from individual mice, the reference compartment was interchangeably defined as CD19⁺ IgD⁺ Fas[−] GL7[−] or CD19⁺ IgD⁺, or CD19⁺ B220⁺ CD138[−] phenotype.

***In vitro* calcium assay.** To examine calcium response of T cells, *in vitro* activated T cells were stained with 2 µM Indo-1 (Invitrogen) in RPMI medium without serum at 37 °C for 30 min, washed twice with cold RPMI containing 1% serum, and then incubated with crosslinking biotinylated anti-CD3 and/or anti-ICOS antibodies as indicated at 4 °C for 30 min. After being washed twice and incubated at 37 °C for 5 min, the baseline Indo-1 fluorescence was collected for 1 min, and then streptavidin was added to a final concentration of 80 µg ml^{−1} before continuous collection of fluorescence signals for additional 5 min.

CD40L mobilization assay of T cells activated *in vivo*. B6 mice were transferred with 2.5 × 10⁶ GFP-expressing OT-II T cells and then immunized with 100 µg NP-OVA intraperitoneally. Seven or eight days later, total splenocytes were isolated, stimulated and stained by a previously described protocol with modifications¹². In brief, cells were incubated with 0.25–1 µg ml^{−1} biotinylated anti-CD3 antibody together with biotinylated anti-ICOS or matching isotype control antibody for 30 min on ice and then washed 3 times with ice-cold PBS before being further incubated with 80 µg ml^{−1} streptavidin in the presence of PE-labelled anti-CD40L or matching isotype control antibody for precisely 3 or 5 min at 37 °C. Cells were

then immediately washed with large volumes of PBS three times, before being either further processed for additional surface staining or directly analysed on a cytometer. An additional group without the initial anti-ICOS incubation but with 1 µM ionomycin added together with the streptavidin was always included. The OT-II T cells were identified by their internal GFP fluorescence, and splenocytes from 2 or 3 mice were pooled for each experiment. To quantitate the effect of ICOS co-stimulation, background-subtracted CD40L MFI values for treated cells were normalized against background-subtracted CD40L MFI for un-stimulated cells to calculate fold changes.

ICOSL measurement of GC and non-GC B cells stimulated *ex vivo*. Total CD19⁺ B cells were isolated by CD19 Microbeads (Miltenyi Biotec) from B6 mice 12 days after NP-KLH immunization. The mixture of GC and non-GC B cells was cultured at 1 × 10⁶ per ml in complete RPMI media supplemented with 10 µg ml^{−1} anti-IgM or 10 µg ml^{−1} anti-CD40 or their isotype-matched control antibodies. Two or eight hours later, cells were recovered, stained for 7-AAD, CD19, IgD, GL7, Fas and ICOSL, and analysed on an LSR II. Non-GC resting B cells were identified as 7-AAD[−] CD19⁺ IgD⁺ GL7^{low} Fas^{low} while GC B cells were identified as 7-AAD[−] CD19⁺ IgD[−] GL7^{high} Fas^{high}. Of note, the apparent fraction of GC B cells in the mixture progressively decreased over the 8-h culture period, consistent with their rapid death rate *in vitro*, while the surviving GC B cells can still be analysed.

ICOSL measurement of GC and non-GC B cells stimulated *in vivo*. B6 mice intraperitoneally immunized with NP-KLH were intravenously injected on day 14 with agonistic anti-CD40 antibody at 200 µg per mouse. Additional B6 mice that were transferred with 2.5 × 10⁶ OT-II T cells and then intraperitoneally immunized with NP-OVA were intravenously given OVA_{323–339} antigenic peptide at 100 µg per mouse. Eight hours after these treatments *in vivo*, mice were killed and splenocytes were stained and analysed as in the *ex vivo* assay described above.

RNA-seq and gene set enrichment analysis (GSEA). B6 mice were transferred each with a combination of 4 × 10⁵ *Icosl*^{−/−} dsRed MD4 B cells, 4 × 10⁵ *Icosl*^{+/+} GFP MD4 B cells, and 10⁵ OT-II T cells and then subcutaneously immunized with HLE-OVA. At 120 h post-immunization, the two types of MD4 GC B cells were sorted from the draining lymph node on an Aria III based on their GFP or dsRed fluorescence combined with CD19, GL7 and Fas expression. A total of 10⁵ sorted cells were subjected to total RNA isolation with the RNeasy Micro kit (Qiagen), and resulting RNA was prepared by the Ovation RNA-Seq System (Nugen) for high throughput sequencing on a Genome Analyzer II. GSEA analysis was carried out using the GSEA software from the Broad Institute^{17,18}. Genes commonly expressed (10,338 genes at reads per kilobase of transcript per million reads mapped > 1) by the two types of GC B cells were ranked by the log2_Ratio_of_Clases metric. Genes that were upregulated by at least threefold following 2-h or 4-h anti-CD40 or anti-IgM stimulation as reported in ref. 19 were used as the test gene sets for analysis of enrichment.

V_H186.2 sequence analysis of NP-specific BMPCs or GC B cells. To analyse V_H sequences of NP-specific BMPCs, single NP-PE⁺ CD138⁺ B220^{low} IgM[−] IgD[−] Gr1[−] bone marrow cells were sorted 21 days after intraperitoneal NP-KLH immunization. Single cells were lysed by incubation at 60 °C for 5 min in the lysis buffer and then reverse-transcribed to complementary DNA using the Superscript cDNA Synthesis Kit (Invitrogen) using the manufacturer's suggested protocol. V_H186.2 fragments were amplified by nested PCR using the following primers (first sense primer: 5'-CTCTTCTTGGCAGCAACAGC-3', first antisense primer: 5'-GCTGCTCAGAGTGTAGAGTTC-3'; second sense primer: 5'-GTGTCCACTCCCAGGTCCAAC-3', second antisense primer: 5'-GTTCCAGGTCACTGTCACTG-3'). PCR products (400 base pairs) were purified by gel electrophoresis and then directly sequenced. Mutations introduced during reverse transcription or PCR are essentially nil due to averaging⁷. To analyse BCR sequences of NP-specific ICOSL^{high} or ICOSL^{low} GC B cells, NP-PE⁺ GC B cells of the desired ICOSL characteristics were sorted from B6 mice that were immunized with NP-KLH 9.5 days previously. Approximately 50 cells were sorted into 1 well of a 96-well plate, and multiple wells of 50-cell sorts from each experiment were processed in parallel by the protocol for single cells as above. The resulting PCR product was cloned into a T vector (Takara), and individual bacterial colonies were picked for sequencing. Identical V_H sequences were counted only once as one clone, and the final result was compiled with unique clones for each category. To analyse BCR sequences of dsRed-expressing wild-type or ICOSL-knockout GC B cells in MD4 recipients, instead of NP-PE, NIP-BSA-biotin (Biosearch Technologies) combined with streptavidin-APC was used to identify antigen-specific B cells.

Intravital imaging of T–B-cell interactions and dynamic distribution patterns. The basic setup and procedure for intravital two-photon imaging of mouse lymph nodes were essentially the same as previously described^{7,16}. To compare duration of contacts between OT-II T cells and *Icosl*^{+/+} or *Icosl*^{−/−} MD4 B cells at the T–B border, the two types of B cells were interchangeably labelled with 2 µM CMFDA or 4 µM CMTPX, and the T cells were labelled with 75 µM CMF₂HC before adoptive transfer. Imaging at a wavelength of 800 nm was conducted 30 to 72 h after HEL-OVA immunization. Contact duration was scored manually. To examine T–B contacts and their dynamic localization behaviours at the GC stage, CFP OT-II T cells, GFP

or dsRed *Icosl*^{+/+} MD4 B cells, and dsRed *Icosl*^{-/-} MD4 B cells were used. To image these cells, the laser was tuned to 850–880 nm to achieve a good compromise for the three fluorescent proteins involved. To capture cell–cell contacts in GCs with high spatial resolution, imaging was conducted using a $\times 25$ water immersion lens (NA = 1.05) at a zoom factor of 3 and a z-step of 1.5 μm . A typical imaging run was 60 min in length. Off-line analyses were conducted using the Imaris software (Bitplane) and ImageJ. Cell–cell contacts were always verified in 3D views. Cell migration was analysed with an Imaris automatic cell tracking module aided with manual verification, and tracks that lasted less than 1 min were excluded from analysis. The cell–cell surface engagement index as defined in Fig. 1 was analysed on z-projection images with manual outline of the cell perimeter and cell–cell contact interface in a double-blinded manner, while measurements were done automatically in ImageJ. To image dynamic distribution patterns of *Icosl*^{+/+} and *Icosl*^{-/-} MD4 GC B cells in relationship with T_{FH} cells, a zoom factor of 1 and a z-step of 3 μm were used. Typical fields of view cover GC volumes rich in T cells. For imaging experiments mentioned above, parameters of adoptive transfer are as follow: (1) imaging of T–B interactions at the T–B border at ~ 30 –72 h post-immunization, 2×10^6 *Icosl*^{+/+} GFP MD4 cells + 2×10^6 *Icosl*^{-/-} dsRed MD4 cells + 10^6 CFP OT-II cells; (2) imaging of GC T–B interactions at 4–5 days post-immunization, 10^5 *Icosl*^{+/+} GFP MD4 cells + 2×10^5 *Icosl*^{-/-} dsRed MD4 cells + 3×10^5 *Icosl*^{-/-} 'dark' MD4 cells + 10^5 CFP OT-II cells; (3) imaging of GC B- and T-cell dynamic distribution at 4–5 days post-immunization, 2×10^4 *Icosl*^{+/+} GFP MD4 cells + 7.8×10^5 *Icosl*^{-/-} or *Icosl*^{+/+} dsRed MD4 cells + 10^5 CFP OT-II cells.

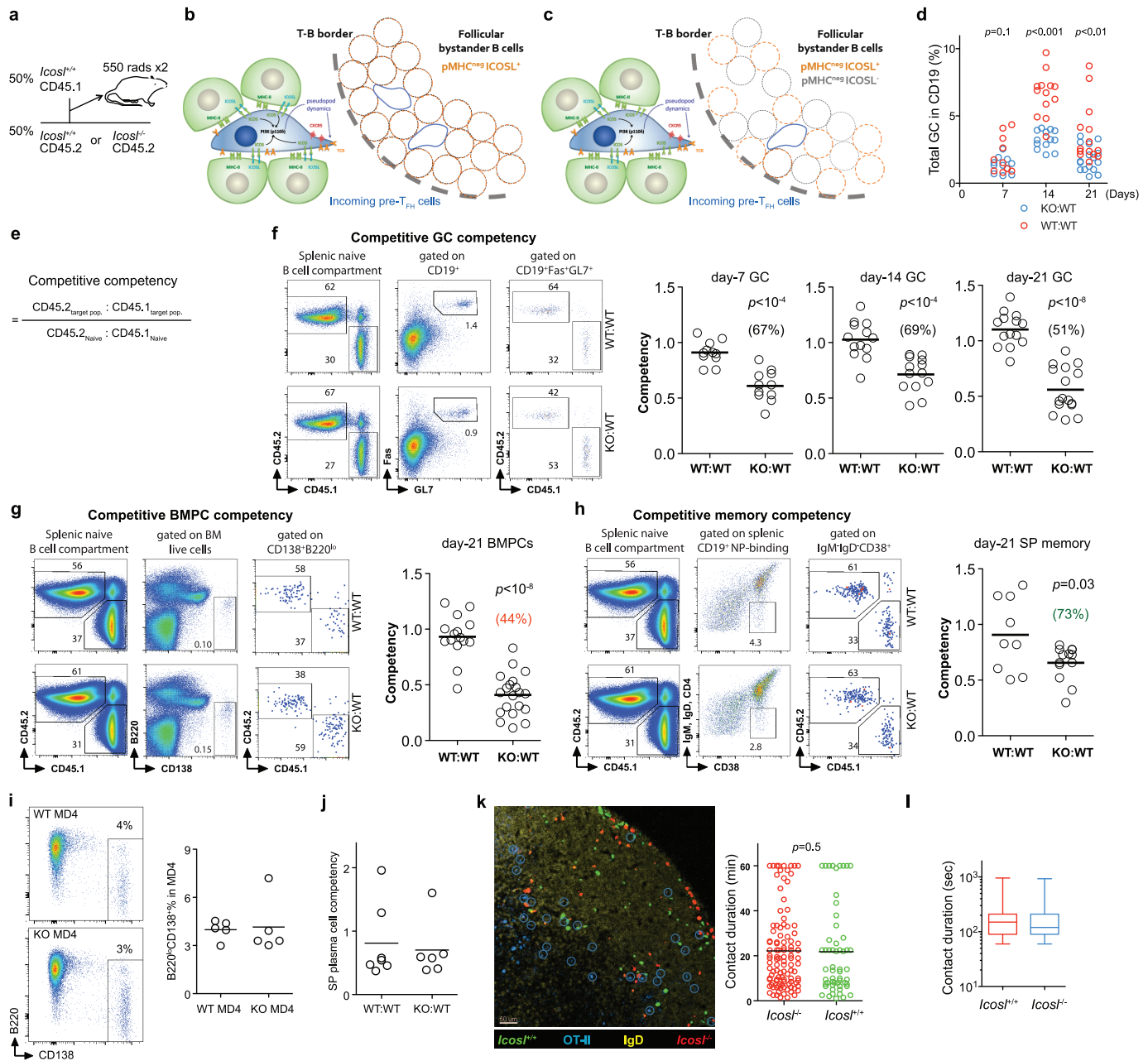
Intravital calcium imaging of T_{FH} cells. The genetically encoded FRET-based calcium reporter YC-Nano50^{CD} was modified from TN-XXL^{CD} (a gift from H. Wekerle)¹⁰ and YC-nano50 (a gift from T. Nagai)⁹. A total of 5×10^5 OT-II T cells transduced with the YC-Nano50^{CD} reporter were transferred into B6 mice together with 5×10^5 dsRed MD4 B cells, without any unlabelled MD4 B cells. Four to five days after immunization with 30 μg HEL-OVA, the draining lymph node was imaged at 830 nm with a zoom factor of 2. To correlate calcium responses of individual T cells with their contacts with B cells, it was necessary to establish a no-contact reference state (R_0) before the contact was analysed. Care was thus taken to image fields that were not over-crowded with GC B cells, as in the apical outer edge of the light zone where individual contacts could be identified and tracked. The typical no-contact reference period of the analysed T cells was 3 to 5 min. See Extended

Data Fig. 3 for additional details on the analytical protocol. In separate experiments, calcium responses of individual T cells were also analysed in the dense core of the light zone. Because T–B contacts here cannot be individually identified or tracked due to the dense packing of GC B cells, the reference state (R_0) of T cells could not be defined in a self-referential, internally controlled manner as above. Instead, we used the average state of all T cells that were outside of the GC and were not in contact with any dsRed MD4 B cells at any time during the imaging session as the reference state, denoted as R_0' for analysing T cells inside the GC LZ during the same imaging session. Technical details of using the R_0' are provided along with quantitative results in Extended Data Fig. 4.

Movie making and playback. Image sequences were annotated in Adobe Photoshop when necessary and rendered as movies with Adobe AfterEffect. Ratiometric images of the calcium imaging were calculated by RatioPlus plugin of ImageJ. All movies are played back at 20 frames per second unless indicated otherwise, and timestamps are all minutes:seconds.

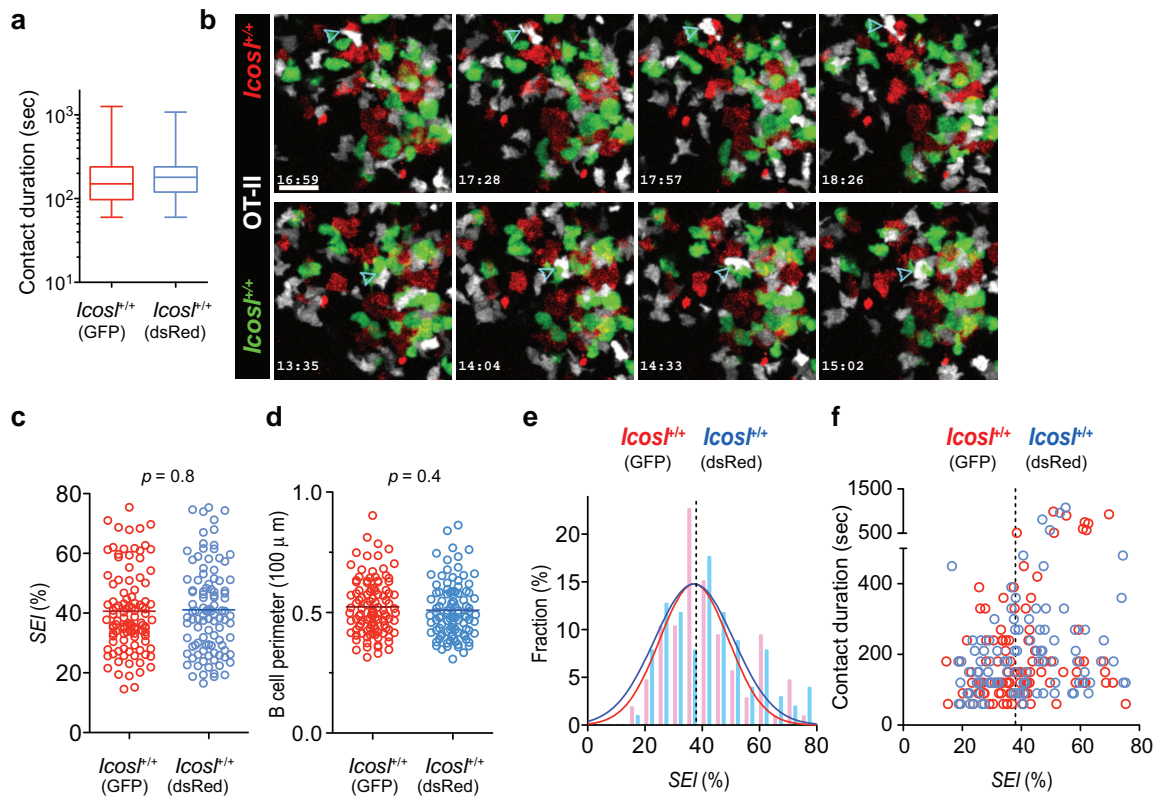
Statistical analysis. Statistical tests with appropriate underlying assumptions on data distribution and variance characteristics were used. Except when noted otherwise, two-tailed Student's *t*-tests were used to compare endpoint means of different groups. Regression and graphing were done with Prism (GraphPad).

15. Weinstein, J. S. *et al.* B cells in T follicular helper cell development and function: separable roles in delivery of ICOS ligand and antigen. *J. Immunol.* **192**, 3166–3179 (2014).
16. Qi, H., Cannons, J. L., Klauschen, F., Schwartzberg, P. L. & Germain, R. N. SAP-controlled T–B cell interactions underlie germinal centre formation. *Nature* **455**, 764–769 (2008).
17. Mootha, V. K. *et al.* PGC-1 α -responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes. *Nature Genet.* **34**, 267–273 (2003).
18. Subramanian, A. *et al.* Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl Acad. Sci. USA* **102**, 15545–15550 (2005).
19. Zhu, X. *et al.* Analysis of the major patterns of B cell gene expression changes in response to short-term stimulation with 33 single ligands. *J. Immunol.* **173**, 7141–7149 (2004).
20. Jacob, J. *et al.* *In situ* studies of the primary immune response to (4-hydroxy-3-nitrophenyl)acetyl. III. The kinetics of V region mutation and selection in germinal center B cells. *J. of Exp. Med.* **178**, 1293–1307 (1993).



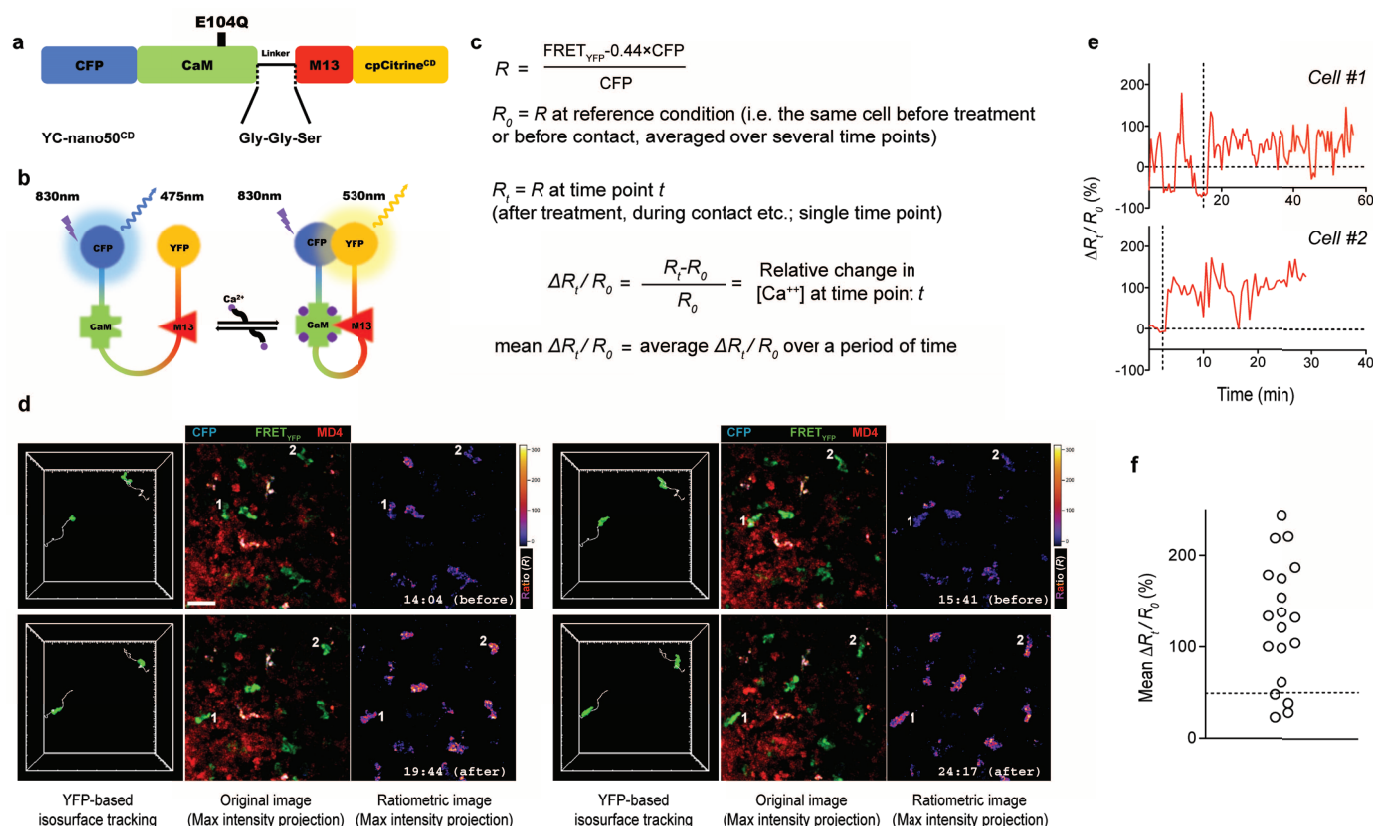
Extended Data Figure 1 | ICOSL-mediated control of GC and BMPC competencies of antigen-specific B cells and the lack of effects of ICOSL-ICOS interactions on cognate T-B contact duration *in vivo*. **a**, Schematic diagram of the protocol of making 50:50 bone marrow (BM) chimaera using CD45.1 wild-type (WT) BM cells mixed with CD45.2 ICOSL wild type or ICOSL-knockout (KO) bone marrow cells. **b**, A schematic diagram depicting the role for follicular non-presenting (pMHC^{neg}), bystander B-cell ensemble in promoting ICOS-driven, PI3K-dependent motility of T cells and their recruitment into the follicle from the T-B border. **c**, A schematic diagram depicting the reduced but non-abrogated effect on follicular recruitment of T cells by the non-presenting bystander B-cell ensemble in the *Icosl*^{+/+}*Icosl*^{-/-} 50:50 mixed bone marrow chimaera. **d**, Percentages of total GC B cells in CD19⁺ cells in chimaeras of indicated types at indicated time points after immunization with NP-KLH. These are the same chimaeric mice analysed below. **e**, Equation for calculating CD45.2 competitive competencies. **f–h**, Pseudo-colour scatter plots show gating strategies and the scatter plots show CD45.2 competency values in individual chimaeric mice for the CD19⁺GL7^{high}Fas^{high} GC compartment at day 7, 14 and 21 post-immunization (**f**), the B220^{low}CD138^{high} BMPC compartment at day 21 (**g**), and the NP-binding, isotype-switched splenic memory compartment at day 21 (**h**) following immunization with 100 µg NP-KLH (examples in **g** and

h from the same mouse assayed for both BMPC and memory B cells). Each symbol represents one mouse. Lines denote the means. Data are pooled from 3 (day-7 GC, day-21 memory), 4 (day-14 GC), or 5 (day-21 BMPC) experiments. The percentages in parentheses were ratios of mean values between knockout:wild-type and wild-type:wild-type. **i**, Percentages of plasma cells developed from *Icosl*^{+/+} or *Icosl*^{-/-} MD4 B cells at day 5 after HEL-OVA immunization in recipient mice that were also given OT-II cells. **j**, Development of NP-specific splenic plasma cells in knockout:wild-type and wild-type:wild-type mixed chimaera 6 days after NP-KLH immunization. The competitive competency is calculated according to **e**. **k**, T-B interactions in the presence or absence of ICOSL on the B-cell partner. Left, the T-B border distribution pattern of CFP OT-II T cells and *Icosl*^{+/+} MD4 GFP B cells and *Icosl*^{-/-} MD4 dsRed B cells 36–72 h post-immunization in the draining lymph node taken after intravital imaging analyses; follicle identified by IgD staining, each CFP⁺ T cell highlighted with a circle; scale bar, 50 µm. Right, contact duration between OT-II T cells and the two types of MD4 B cells as measured by intravital imaging (See also Supplementary Video 1a). **l**, Interactions between CFP OT-II T cells and GFP *Icosl*^{+/+} or dsRed *Icosl*^{-/-} MD4 B cells visualized 96–120 h after HEL-OVA immunization. Duration of T-B contacts involving 114 *Icosl*^{+/+} and 102 *Icosl*^{-/-} events pooled from 2 experiments. See Supplementary Video 1b and corresponding contact area analyses in Fig. 1.



Extended Data Figure 2 | Comparable T_{FH} entanglement with GFP and dsRed MD4 B cells. To rule out potential effects of GFP and dsRed transgenes on T–B contact, CFP OT-II T cells and co-transferred GFP *Icosl*^{+/+} and dsRed *Icosl*^{+/+} MD4 B cells were visualized 96–120 h after HEL-OVA immunization precisely as done in Fig. 1a–f. **a**, Duration of contacts involving 106 GFP and 102 dsRed *Icosl*^{+/+} MD4 B cells pooled from 2 experiments. **b**, Time-lapse images showing 'entangled' behaviours of OT-II T cells (arrowheads)

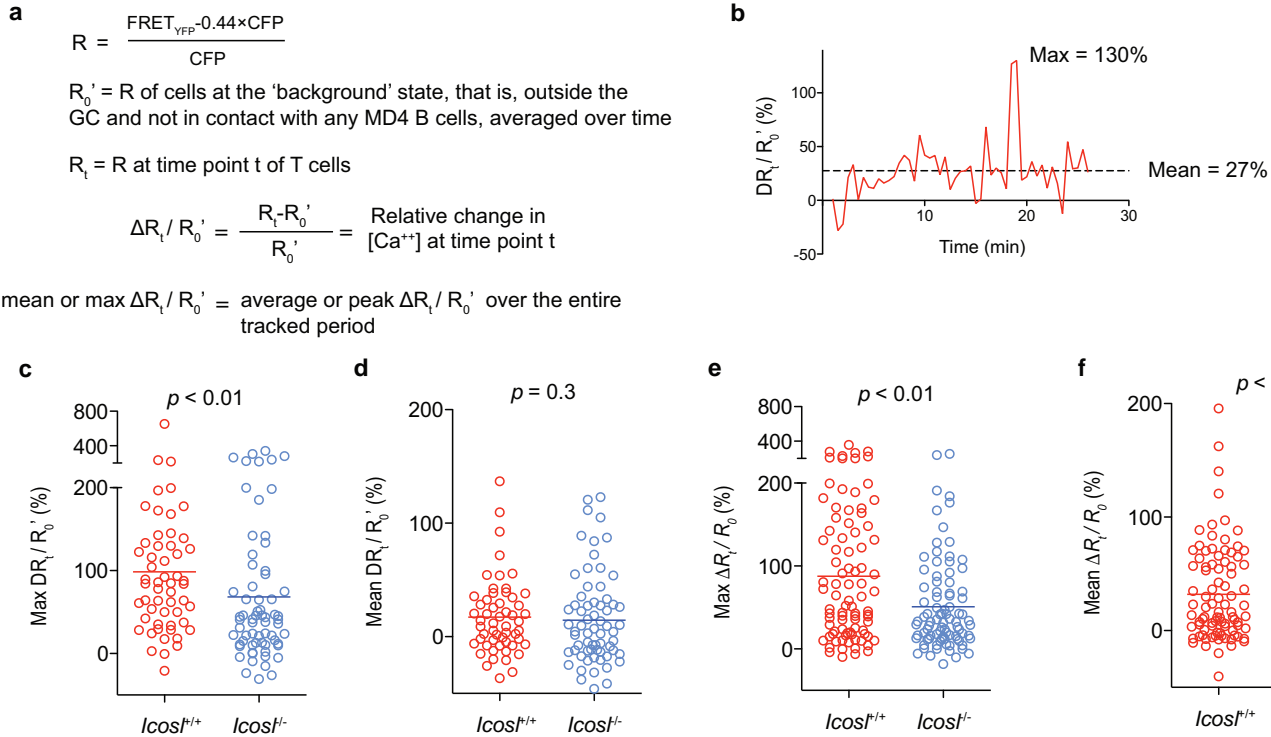
contacting the two types of B cells; time is shown as min:s. See corresponding Supplementary Video 3. Scale bar, 20 μm. **c**, Scatter plots of the SEI for 106 GFP and 102 dsRed MD4 B cells. **d**, Perimeters of analysed B cells of the two types. **e**, **f**, SEI histograms and fitted Gaussian curves (R^2 for goodness of fit: GFP, 0.79; dsRed, 0.79) (**e**) and the plot of SEI against contact duration (**f**); the dotted line indicates 38%.



Extended Data Figure 3 | FRET-based calcium imaging of T_{FH} cells *in vivo*.

a, Domain composition of the YC-Nano-50^{CD} calcium reporter, containing the calcium-binding module (CaM-linker-M13) from YC-Nano-50 from ref. 9, and the FRET pair from the codon-diversified CFP and cpCitrine described in ref. 10. **b**, A diagram of the calcium-sensing FRET process. The normalized 2-photon excitation efficiency of CFP and YFP at 830 nm is ~80% and ~5%, respectively. **c**, Formulas for quantitative parameters. **d–f**, Demonstration of analysing calcium fluxes of T_{FH} cells in response to antigen *in vivo*. MD4 GCs were induced and imaged as described in Methods. During the imaging session, 100 μg OVA_{323–339} peptide was intravenously injected to acutely trigger the OT-II T cells. **d**, Image sequences of four time points, two before and two after the OVA injection. For each time point, middle: fluorescence overlay; scale bar, 50 μm; left, two representative T cells tracked by iso-surfaces built on YFP fluorescence; right, pseudo-colour FRET ratio image (R image, as defined in **c**)

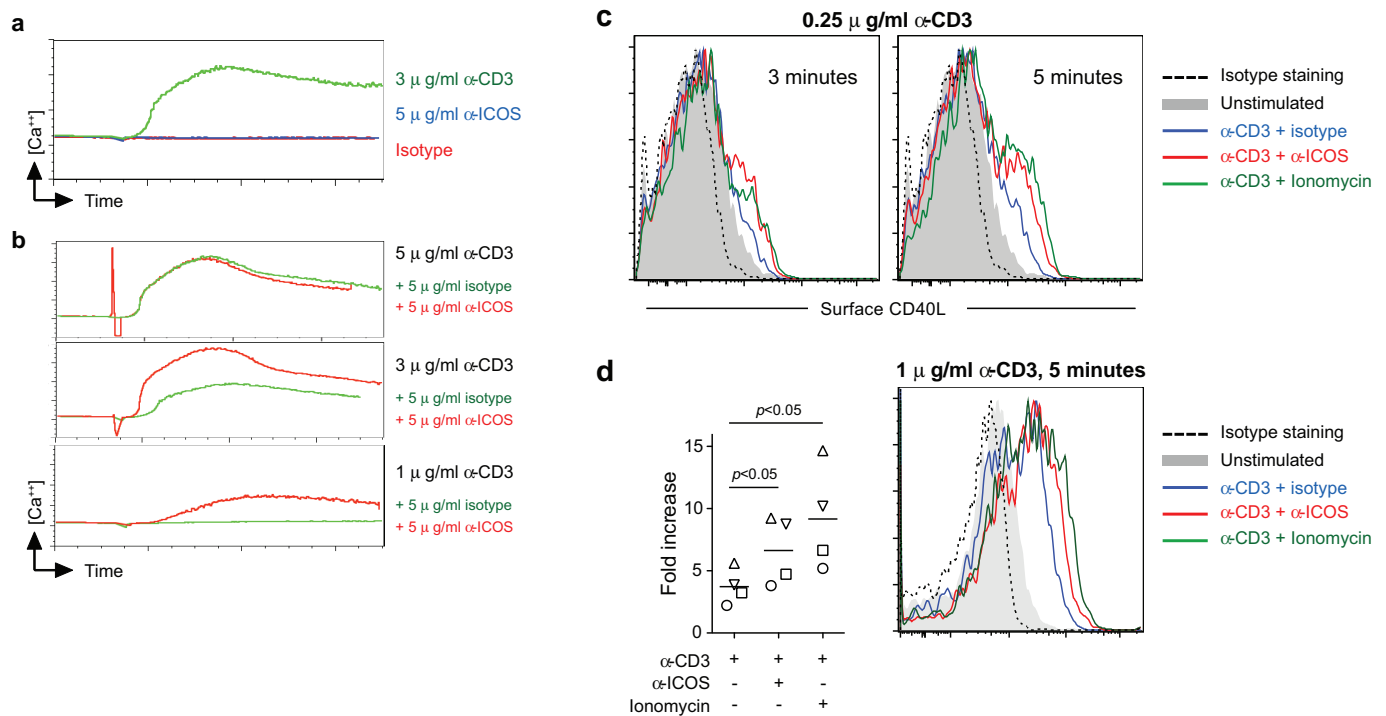
calculated pixel by pixel by the RatioPlus plug-in of ImageJ. Note cell no. 1 is within the GC proper in the lower left corner, while cell no. 2 is in the area immediately surrounding the GC. **e**, $\Delta R_t / R_0$ calculated according to **c** for the two T cells tracked in **d**. Dotted lines indicate the time of OVA injection. R_0 here is the average R of all trackable time points before OVA injection. Total fluorescence intensity signals from voxels encapsulated within the iso-surface (cell) were used to calculate R_0 and R_t . Fluctuations of $\Delta R_t / R_0$ seen in cell no. 1 before OVA injection likely reflect triggering by endogenous B cells in the GC. **f**, Mean $\Delta R_t / R_0$ averaged over the first 5 min after OVA injection for 19 OT-II T cells imaged in 2 independent experiments. The dotted line indicates an average 50% increase in R over R_0 . The mean $\Delta R_t / R_0$ is an internally referenced parameter and can be used to compare cells imaged in different sessions.



Extended Data Figure 4 | Calcium imaging of T_{FH} cells in the core LZ densely populated with MD4 B cells. Intravital imaging of YC-Nano-50^{CD}-transduced OT-II T cells was done as in Fig. 2h–l except with imaging fields covering the core light zone and a change of reference state definition.

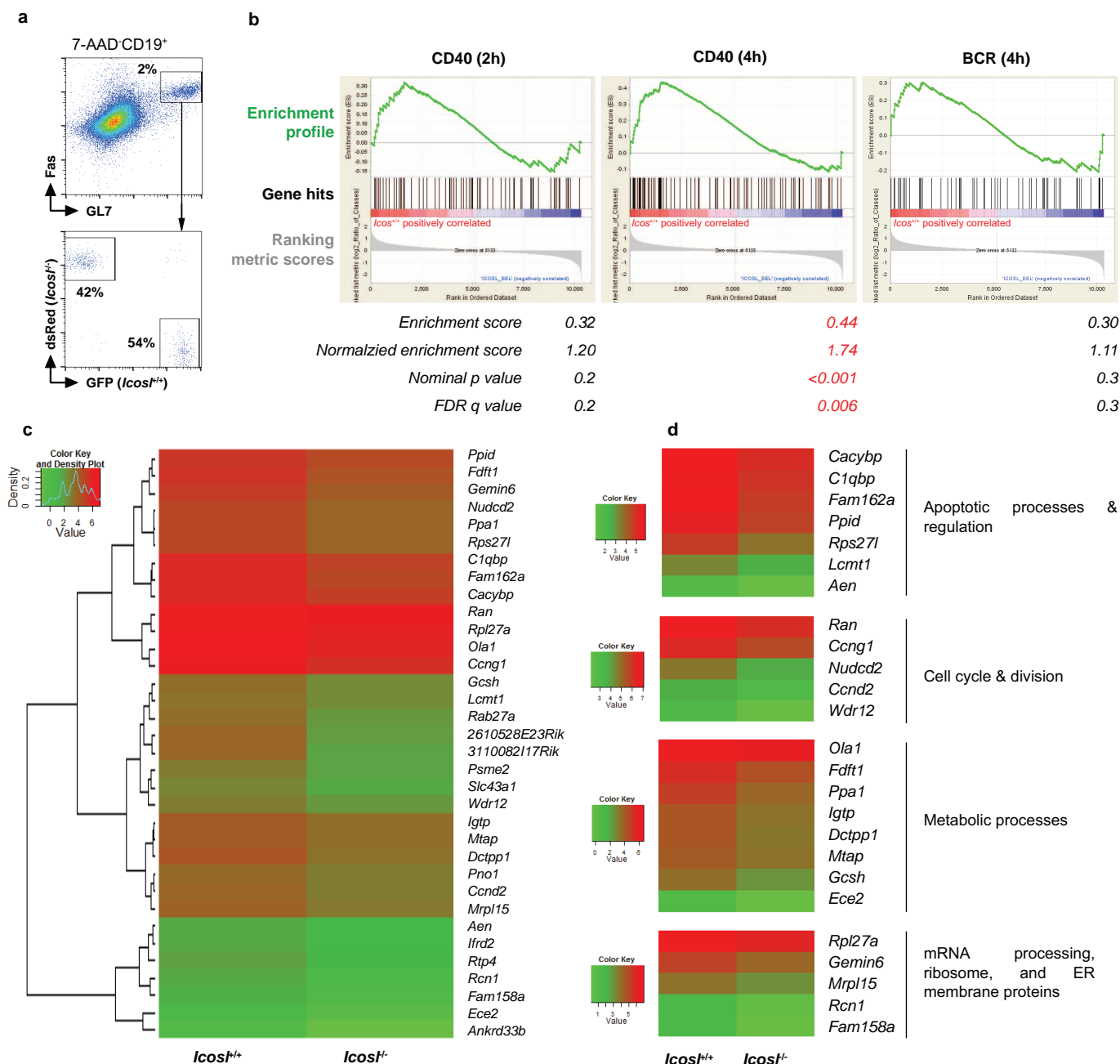
a, Formulas for quantitative parameters. Of note, R_0' , as an ensemble average of all T cells outside GCs without ever contacting MD4 B cells in the imaging session, is theoretically equivalent to the average of R_0 s of all T cells quantitated in Fig. 2. **b**, A $\Delta R_t/R_0'$ trace to illustrate the sensitivity of maximum $\Delta R_t/R_0'$ quantitation and the insensitivity of mean $\Delta R_t/R_0'$ quantitation due to averaging of infrequent high-amplitude calcium spikes. **c**, **d**, Maximum (**c**) or

mean (**d**) $\Delta R_t/R_0'$ of 55 OT-II T cells visualized in $Icosl^{+/+}$ GCs and 67 in $Icosl^{-/-}$ GCs. Data were pooled from three experiments. **e**, **f**, For comparison, maximum (**e**) or mean (**f**) $\Delta R_t/R_0'$ of all 88 OT-II T cells contacting $Icosl^{+/+}$ MD4 GC B cells and 88 contacting $Icosl^{-/-}$ cells that were analysed in Fig. 2h–l are shown. Note comparable sensitivities of **c** and **e** in detecting differences between T cells under the two conditions. The apparent failure of **d** to detect differences between the two conditions is consistent with prevailing short GC T–B contacts and infrequent high-amplitude calcium spikes averaged out by relatively idle periods.

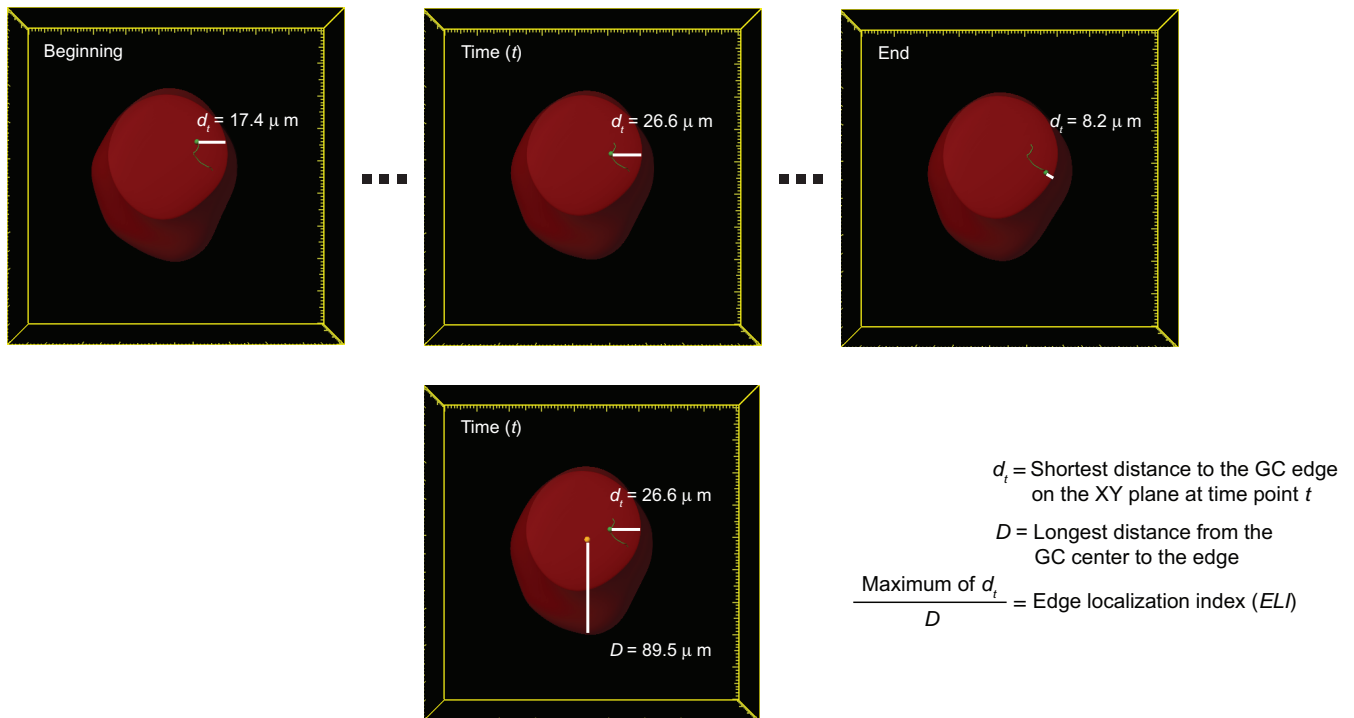


Extended Data Figure 5 | ICOS co-stimulates calcium signalling and CD40L mobilization. **a, b**, Calcium fluxes detected by Indo-1 in T-cell blasts that were incubated on ice with biotinylated anti-CD3 or anti-ICOS or isotype control antibodies separately (**a**) or incubated with the indicated mixture of anti-CD3 anti-ICOS or isotype control antibodies on ice (**b**) before streptavidin stimulation at 37 °C. Data represent three independent experiments. **c**, Surface staining of CD40L following stimulation for 3 (left) or 5 (right) minutes with 0.25 $\mu\text{g ml}^{-1}$ anti-CD3 plus 5 $\mu\text{g ml}^{-1}$ anti-ICOS or isotype or

1 μM ionomycin according to the mobilization assay protocol detailed in the Methods. One of two experiments with similar results is shown. **d**, Quantitation of four independent ICOS co-stimulated CD40L mobilization assays (left, each matched symbol denoting one experiment) and the histogram overlay from one representative experiment (right). Background-subtracted CD40L MFI on un-stimulated cells is defined as unit 1. The *P* values were calculated with paired *t*-tests.



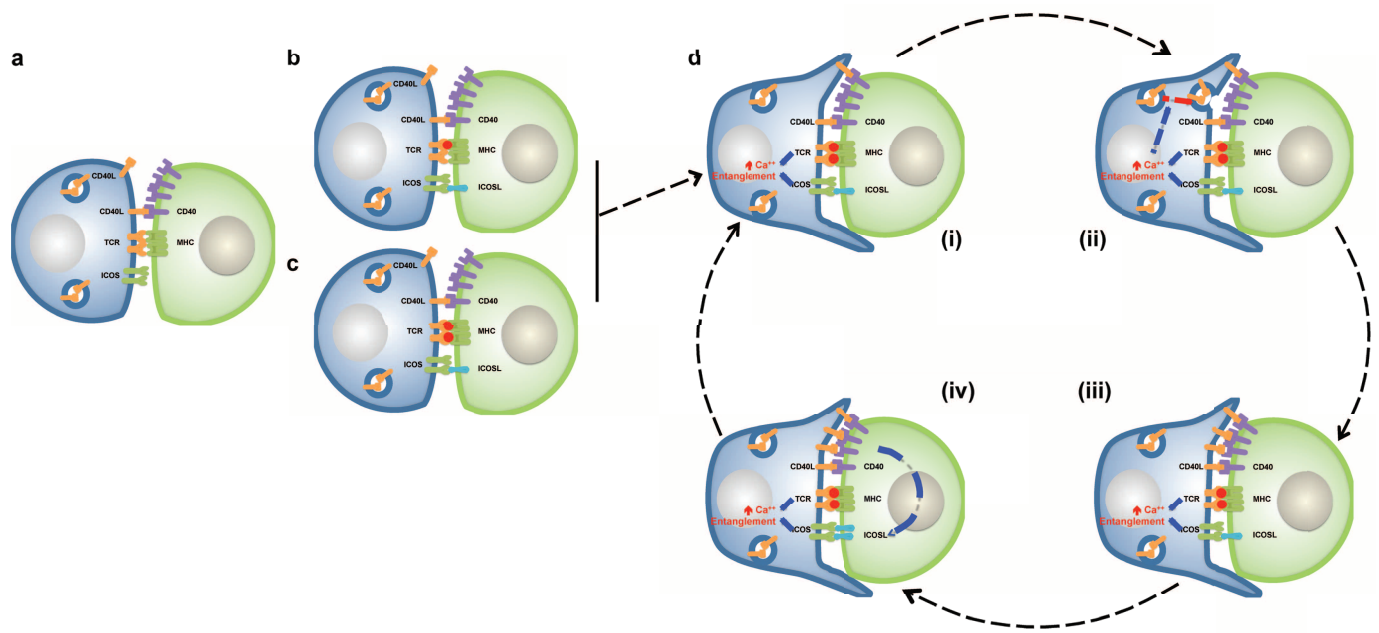
(left) or 4-h (middle) anti-CD40 or anti-IgM (right) stimulation as reported in ref. 18. Compared to *Icosl*^{+/-} GC B cells, *Icosl*^{+/+} cells *in vivo* are strongly imprinted with the transcription signature that is significantly upregulated by prolonged CD40 signalling. **c**, **d**, Genes in the core enrichment for the CD40 (4 h) group, presented as a cluster-analysed heatmap of expression levels (log₂-transformed RPKM) of the two types of cells (c) or as subclasses manually categorized according to Gene Ontology (<http://www.geneontology.org>) (d). Nine genes that were either unclassifiable or with unknown functions were omitted.



Extended Data Figure 7 | Calculation of the edge localization index (ELI).

To analyse imaging data presented in Fig. 3, the GC volume is detected and marked by an iso-surface using dsRed fluorescence of the competitor GC cell population. Individual T or B cells within the GC are tracked, as represented by the green dot and line trace here. At any time point t , the shortest distance of the

cell to the GC iso-surface is calculated by a custom Matlab script as d_t . The maximal value of d_t for a given cell represents its longest distance away from the GC edge. The maximal radius of the GC on a xy plane is defined as D . The maximum of d_t for a given cell divided by D is the edge localization index of the cell. The closer ELI is to 0, the closer the cell is located to the GC edge.



Extended Data Figure 8 | A schematic representation of entanglement and ICOSL-driven feed-forward GC selection. **a–d**, Normally T_{FH} cells display a low level of CD40L on the cell surface, which can in principle be used by any GC B cells that constitutively express the CD40 receptor, regardless of ICOSL expression (**a**); GC B cells carrying BCR of the same specificity but different affinities present different amounts of peptide–MHC (major histocompatibility complex) complexes (red dot: peptide epitope) (**b**, **c**), and utilize the intercellular feed-forward loop (**d**) differently. More intense peptide–MHC antigen presentation coupled with ICOSL costimulation leads to enhanced calcium response and better entanglement by the T cell (**i**); the calcium

response triggers rapid CD40L externalization from the intracellular store (**ii**), and far more CD40L becomes available and used by the presenting B cell (**iii**); CD40L signals and probably additional help signals from the T cells promote ICOSL surface expression (**iv**), which prepare the B cells for more efficient entanglement with the T cells. Note that one B cell does not need to repeatedly entangle with the same T cells, and there must be a way by which ICOSL is down-modulated even on outcompeting GC B cells. ICOSL downregulation by BCR signalling as suggested in Fig. 3 could be a potential mechanism (not depicted).

Extended Data Table 1 | V_H mutations in NP-binding GCs at day 9.5 in mixed chimaera

Genotype (number of clones sequenced)	V _H 186.2 usage	V _H 186.2 mutated	Mutations (mean±SEM) per V _H 186.2 gene ^a		W33L in V _H 186.2-carrying clones ^b		Counter- Selection Index ^c
CD45.2 <i>Icosl</i> ^{-/-} (37)	37/37	37/37	3.5±0.2	<i>p</i> <0.001	23/37 (62%)	<i>p</i> =0.6	<div><div>62%/69%</div><div>36%/78%</div><div>= 1.9</div></div>
CD45.1 <i>Icosl</i> ^{+/+} (32)	32/32	31/32	2.3±0.2		22/32 (69%)		
CD45.2 <i>Icosl</i> ^{+/+} (47)	47/47	43/47	2.7±0.3	<i>p</i> =0.4	17/47 (36%)	<i>p</i> <0.001	
CD45.1 <i>Icosl</i> ^{+/+} (36)	36/36	35/36	3.1±0.3		28/36 (78%)		

^aMann-Whitney non-parametric test for mutation rates. ^bFisher's exact test for W33L frequencies between the two cell types within the same chimaeras. ^cThe counter-selection index as measured by ratios of W33L frequencies reflects the selective pressure that ICOSL-knockout B cells experience after normalization of inherent difference between CD45.1 and CD45.2 cells. Unit 1 means neutral (or at the competition equilibrium), while values greater than 1 indicate a higher pressure and more stringent selection against ICOSL-knockout B cells. It is currently unclear why CD45.2 wild-type cells harbour significantly fewer W33L mutants than CD45.1 counterparts in the wild-type:wild-type chimaera (17/47 versus 28/36). CD45.2 and CD45.1 GC B cells were sorted from chimaeric mice of indicated types on day 9.5 post NP-KLH immunization and subjected to V_H sequence analyses. See Supplementary Notes 3 and 4 for additional details on appropriate interpretation of these and related results.

Extended Data Table 2 | V_H mutations in NP-binding BMPCs at day 21 in mixed chimaera

Genotype (number of clones sequenced)	V _H 186.2 usage	V _H 186.2 mutated	Mutations (mean±SEM) per V _H 186.2 gene ^a		W33L in V _H 186.2-carrying clones ^b		Counter-Selection Index _c
CD45.2 <i>Icosl</i> ^{-/-} (46)	46/46	40/46	2.5±0.2	<i>p</i> <0.05	33/46 (72%)	<i>p</i> <0.05	<div><div>72%/46%</div><div>44%/69%</div><div>= 2.5</div></div>
CD45.1 <i>Icosl</i> ^{+/+} (39)	35/39	24/35	1.7±0.3		16/35 (46%)		
CD45.2 <i>Icosl</i> ^{+/+} (35)	34/35	26/34	2.1±0.4	<i>p</i> =0.7	15/34 (44%)	<i>p</i> =0.05	
CD45.1 <i>Icosl</i> ^{+/+} (33)	32/33	29/32	2.3±0.3		22/32 (69%)		

^aMann-Whitney non-parametric test for mutation rates. ^bFisher's exact test for W33L frequencies between the two cell types within the same chimaeras. ^cThe counter-selection index as measured by ratios of W33L frequencies reflects the selective pressure that ICOSL-knockout B cells experience after normalization of inherent difference between CD45.1 and CD45.2 cells. Unit 1 means neutral (or at the competition equilibrium), while values greater than 1 indicate a higher pressure and more stringent selection against ICOSL-knockout B cells. CD45.2 and CD45.1 BMPCs were sorted from chimaeric mice of indicated types on day 21 post NP-KLH immunization and subjected to V_H sequence analyses. The non-V_H186.2 sequences recovered were V_H186.2 analogues of the J558 V_H family, which were previously observed as frequent in early NP-specific GCs but outcompeted by 186.2 clones by the time of peak GC response²⁰. Data in this table were generated in series of chimaera experiments conducted 2 years before those in Extended Data Table 1. See Supplementary Notes 3 and 4 for additional details on appropriate interpretation of these and related results. The CD45.2 NP-binding BMPC competency, calculated as in Extended Data Fig. 1, is 0.44 ± 0.1 and 1.01 ± 0.1 in the *Icosl*^{-/-}:*Icosl*^{+/+} and *Icosl*^{+/+}:*Icosl*^{+/+} chimaeras, respectively (mean ± s.e.m. of 4 independent experiments, *P* < 0.01).

EFF-1-mediated regenerative axonal fusion requires components of the apoptotic pathway

Brent Neumann¹, Sean Coakley^{1*}, Rosina Giordano-Santini^{1*}, Casey Linton¹, Eui Seung Lee², Akihisa Nakagawa², Ding Xue² & Massimo A. Hilliard¹

Functional regeneration after nervous system injury requires transected axons to reconnect with their original target tissue. Axonal fusion, a spontaneous regenerative mechanism identified in several species, provides an efficient means of achieving target reconnection as a regrowing axon is able to contact and fuse with its own separated axon fragment, thereby re-establishing the original axonal tract^{1–7}. Here we report a molecular characterization of this process in *Caenorhabditis elegans*, revealing dynamic changes in the subcellular localization of the EFF-1 fusogen after axotomy, and establishing phosphatidylserine (PS) and the PS receptor (PSR-1) as critical components for axonal fusion. PSR-1 functions cell-autonomously in the regrowing neuron and, instead of acting in its canonical signalling pathway⁸, acts in a parallel phagocytic pathway that includes the transthyretin protein TTR-52, as well as CED-7, NRF-5 and CED-6 (refs 9–12). We show that TTR-52 binds to PS exposed on the injured axon, and can restore fusion several hours after injury. We propose that PS functions as a ‘save-me’ signal for the distal fragment, allowing conserved apoptotic cell clearance molecules to function in re-establishing axonal integrity during regeneration of the nervous system.

Axonal fusion, which occurs spontaneously in several invertebrate species^{1–7}, is a highly efficient means to re-establish the connection between an injured neuron and its target tissue; the proximal axon which is still attached to the cell body regrows towards, reconnects, and fuses with its separated distal fragment. We and others have shown that axonal fusion occurs in *C. elegans* mechanosensory neurons^{4,7}. However, the genetic components of this process remain largely unidentified, with the nematode-specific fusogen, epithelial fusion failure 1 (EFF-1), the only protein known to be involved⁴.

We expressed green fluorescent protein (GFP) specifically within the six mechanosensory neurons of *C. elegans* (Fig. 1a) and performed laser axotomy of the posterior lateral mechanosensory (PLM) neurons. We then analysed severed axons that re-established a contact between the regrowing proximal axon and its separated distal fragment (Extended Data Fig. 1); maintenance, or inhibition of degeneration of the distal axonal fragment, was used as evidence of successful fusion, whereas degeneration indicated unsuccessful fusion⁷. In wild-type animals, 80% of axons that displayed proximal–distal reconnection underwent successful fusion (Fig. 1b, d). EFF-1, a trimeric fusogen similar to class-II viral fusion proteins^{13,14}, mediates a wide range of fusion events in *C. elegans*^{15,16}, including the sculpting of dendritic arbors through self-fusion of excessive branches¹⁷. Animals lacking EFF-1 displayed a strong defect in axonal fusion⁴ (Fig. 1c, d). To determine where EFF-1 is required, we expressed a GFP-tagged version of this protein specifically in the mechanosensory neurons. This transgene could rescue the axonal fusion defect of *eff-1* null animals, indicating a cell-autonomous function of EFF-1 in PLM (Fig. 1d). Prior to axotomy, we observed EFF-1::GFP in an indiscriminate punctate pattern throughout the PLM cell body and axon (Fig. 1e and Extended Data Fig. 2a). Soon after axotomy, a ‘cap’ of EFF-1::GFP appeared on the membrane at the tips of both the proximal and distal

segments by the cut site (Extended Data Fig. 2b, c). Following regenerative growth, we consistently observed EFF-1::GFP on the membrane of the growth cone (Fig. 1f and Extended Data Figs 2d and 3a, b), suggesting rapid re-localization of EFF-1 to mediate fusion once contact is re-established.

We speculated that axonal fusion might share similarities with apoptotic cell corpse clearance, in which plasma membrane phospholipid asymmetry is lost. PS, normally restricted to the cytoplasmic leaflet, becomes externalized to the exoplasmic leaflet of the apoptotic cell, which enables its recognition by phagocytic cells¹⁸. Two different loss-of-function deletions in the PS receptor gene *psr-1* (refs 8 and 19) caused a strong reduction in axonal fusion (Fig. 2a, b). Furthermore, double mutants between *eff-1* and *psr-1* were no more defective than *eff-1* mutants (Fig. 2b), indicating that *psr-1* and *eff-1* are in the same genetic pathway. PSR-1 contains two highly conserved functional domains, an extracellular lysine-rich PS binding domain²⁰, and an intracellular JmjC domain with multiple enzymatic activities^{21–23}. We expressed different versions of full-length *psr-1* cDNA selectively in the mechanosensory neurons of *psr-1* mutant animals. Expression of wild-type cDNA rescued axonal fusion (Fig. 2c), demonstrating that PSR-1 functions cell-autonomously within PLM. Next, we expressed *psr-1* cDNA carrying mutations that disrupt either its PS binding activity (K308E/K315E)²⁰ or its Fe(II) binding site (H192A/D194A) in the JmjC domain²⁰ and found that they no longer rescued the axonal fusion defect (Fig. 2c). Thus, PSR-1 requires both the extracellular PS binding domain and the intracellular JmjC domain for its cell-autonomous activity in axonal fusion.

To determine where PSR-1 functions within PLM, we analysed the localization of a functional mCherry-tagged version of PSR-1 in this neuron (Fig. 2c). PSR-1 localized to the nucleus and mitochondria of PLM (Fig. 2d and Extended Data Fig. 4a, b). However, following axotomy, we observed an accumulation of PSR-1::mCherry towards the end of the regrowing proximal axon (Fig. 2d and Extended Data Fig. 4b). Thus, PSR-1 is normally confined to the nucleus and mitochondria of PLM, but may be re-localized following injury.

PSR-1 functions in a signalling pathway that includes CED-2/CrkII, CED-5/Dock180, CED-12/ELMO and CED-10/RAC1 to mediate the recognition and removal of apoptotic cells^{8,24–26} (Fig. 3a). We found that other components of this engulfment pathway are not required for axonal fusion, as animals carrying mutations in *ced-2*, *ced-5*, *ced-12*, or *ced-10* were not defective (Extended Data Fig. 5a). Similarly, the core apoptotic pathway is not required, as mutations in CED-9/BCL-2, CED-4/APAF-1, or cell-killing caspase *ced-3* did not significantly alter axonal fusion (Supplementary Table 1). Next, we analysed components of another, partly redundant cell corpse engulfment pathway, which includes the secreted PS binding protein TTR-52/transsthyretin¹⁰, the lipid-binding protein NRF-5 (ref. 12), the membrane-bound CED-7/ABC transporter¹¹, the transmembrane receptor CED-1/LRP1/MEGF10 (ref. 27), and the intracellular adaptor CED-6/GULP⁹ (Fig. 3a). We found that TTR-52 was critical for successful axonal fusion (Fig. 3b and Extended Data Fig. 5b).

¹CJCADR, Queensland Brain Institute, The University of Queensland, Brisbane QLD 4072, Australia. ²Department of Molecular, Cellular, and Developmental Biology, University of Colorado, Boulder, Colorado 80309, USA.

*These authors contributed equally to this work.

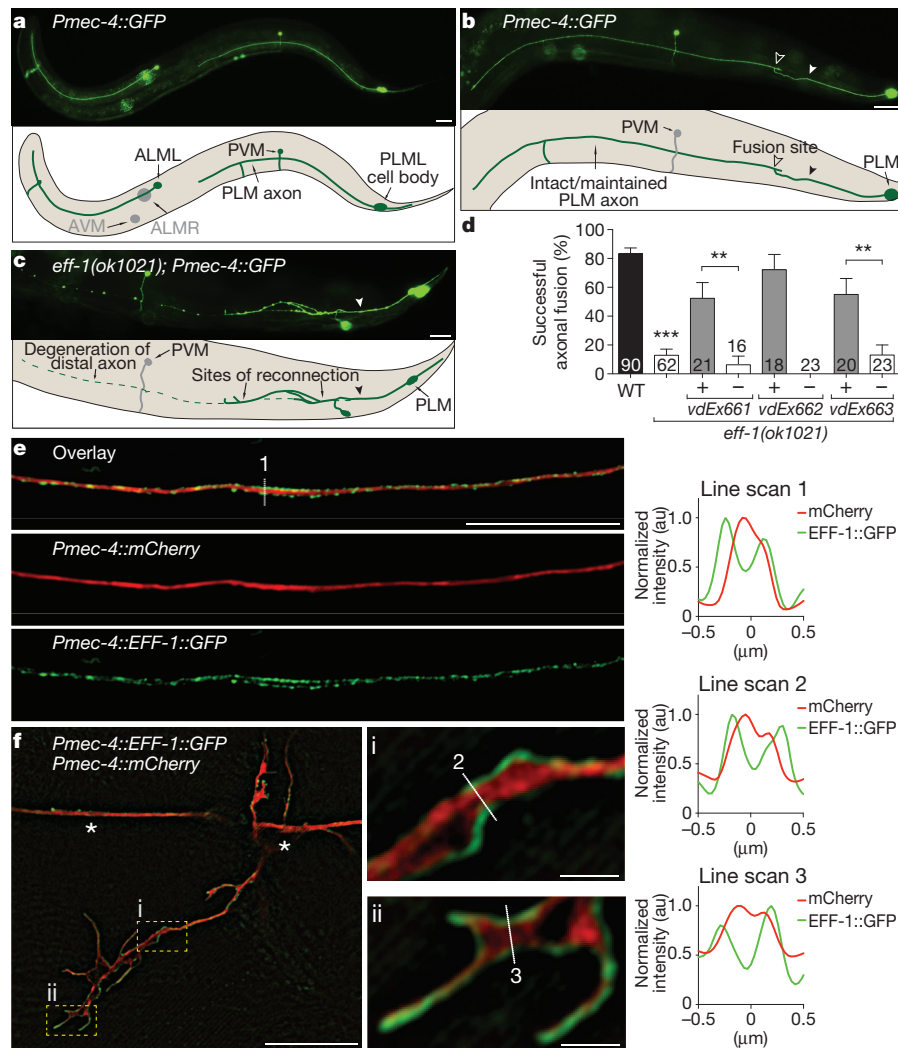


Figure 1 | EFF-1 mediates axonal fusion cell-autonomously and localizes to the membrane of growth cones. **a**, A wild-type (WT) *zdis5*(*Pmec-4::GFP*) animal with fluorescent mechanosensory neurons. Anterior is left and ventral is down in this and all following images. Out of focus neurons are shown in grey in the schematic. Scale bar, 25 μ m. Image representative of 242 animals. **b**, Successful axonal fusion in a WT animal, 24 h post-axotomy. Filled arrowhead points to the cut site; open arrowheads point to the fusion site. Scale bar, 25 μ m. Image representative of 75 animals. **c**, Defective axonal fusion in an *eff-1(ok1021)* mutant animal 48 h post-axotomy. Filled arrowhead points to the cut site. Scale bar, 25 μ m. Image representative of 54 animals. **d**, Quantification of axonal fusion in *eff-1(ok1021)* animals compared to WT, and cell-autonomous rescue with expression of *Pmec-4::EFF-1::GFP* in independent transgenic lines. Error bars indicate standard error of proportion.

As TTR-52 is normally secreted from intestinal cells to bind PS exposed by apoptotic cells¹⁰, we asked whether TTR-52 binding to PS also occurred after axonal injury. TTR-52 tagged with mCherry and expressed from a heat-shock promoter rescued the axonal fusion defect when induced either at the time of injury, or up to 6 h after injury (Fig. 3b and Extended Data Fig. 5c–g). We then used the same functional transgene to visualize TTR-52 before and after PLM axotomy. Following axotomy, TTR-52 rapidly localized to both the distal and proximal axon segments, with slightly stronger mCherry signals observed on the distal segment (Fig. 3d, e). These results suggest that TTR-52 functions as a secreted molecule that binds injured axonal fragments to mediate their regenerative fusion. To further confirm that TTR-52 was binding PS, we analysed the localization of a secreted version of Annexin V (sAnxV), a highly efficient and specific calcium-dependent PS binding protein²⁸. PS was rapidly exposed on the proximal and distal PLM axon segments following

axotomy, with the sAnxV::mRFP signal strongest on the distal segment (Fig. 3f and Extended Data Fig. 5h). This localization pattern closely matched that of TTR-52::mCherry (Fig. 3d, e), supporting a model in which PS is exposed on the axonal membrane after injury and acts as a ‘save-me’ signal for TTR-52 binding, thereby triggering reconnection and fusion.

Next we examined the involvement of the other components of the TTR-52 engulfment pathway. Inactivation of *ced-1* caused a modest defect, whereas animals lacking NRF-5, CED-7, or CED-6 displayed strong fusion defects (Fig. 3c and Supplementary Table 1). Double mutants between *psr-1* and either *ttr-52*, *ced-1*, *nrf-5*, *ced-7* or *ced-6* were no more defective than single mutant animals (Fig. 3c and Supplementary Table 1), indicating that PSR-1 functions within this TTR-52-mediated pathway during axonal fusion. Similar to *psr-1*; *eff-1* double mutant animals (Fig. 2b), double mutants between *eff-1* and either *ttr-52*, *ced-1*,

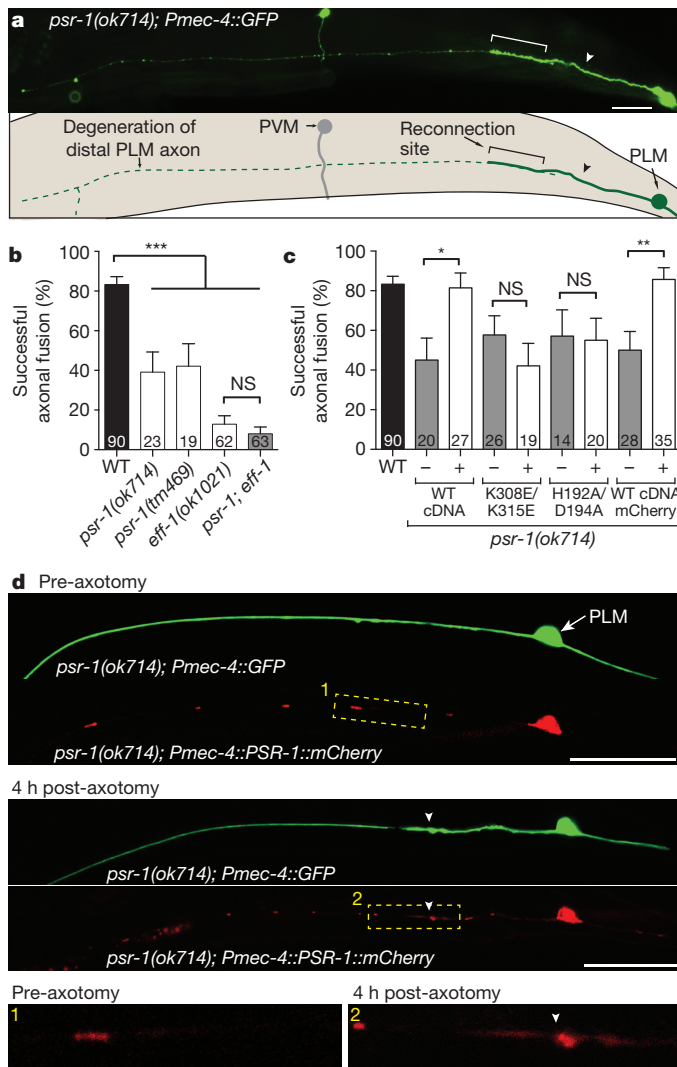


Figure 2 | PSR-1 functions cell-autonomously in regulating axonal fusion. **a**, Unsuccessful fusion in a *psr-1(ok714)* mutant animal 24 h after axotomy. Bracket highlights the region of reconnection. Scale bar, 25 μ m. Image representative of 14 animals. **b**, Quantification of successful fusion events in wild-type (WT, black bar), and single (white bars) and double (grey bar) mutant animals. **c**, Cell-autonomous rescue in *psr-1(ok714)* animals with expression of WT *psr-1* cDNA (untagged or tagged with mCherry) in the mechanosensory neurons. Expression of *psr-1* cDNA carrying mutations in the PS binding (K308E/K315E), or in the JmjC domain (H192A/D194A), fails to rescue the defect; representative results from the six independent lines tested. For **b** and **c**, error bars indicate standard error of proportion. *n* values within each bar. *P* values from *t*-test: **P* < 0.05, ***P* < 0.01, ****P* < 0.001 compared to WT unless marked on graph. NS, not significant. **d**, Single plane confocal images showing the localization of PSR-1::mCherry in PLM in a *psr-1(ok714)* mutant animal before axotomy (top two panels), and 4 h post-axotomy (middle two panels). Bottom panels show magnified images of boxed regions to highlight accumulation of PSR-1 at the end of the regrowing proximal axon. Arrowheads point to cut site. Scale bar, 25 μ m. Image representative of 21 animals.

ced-6 or *ced-7* were no more defective than single *eff-1* mutants (Extended Data Fig. 3c), indicating that these genes act in the same genetic pathway. Importantly, overexpression of EFF-1 specifically in the mechanosensory neurons could overcome the axonal fusion defects caused by mutation of either *psr-1* or *ttr-52* (Extended Data Fig. 3d), demonstrating that EFF-1 functions genetically downstream of these molecules. Notably, although mutations in some axonal fusion genes caused defective PLM regrowth, we found no correlation between regrowth and axonal fusion (Extended Data Fig. 6 and Supplementary Table 2).

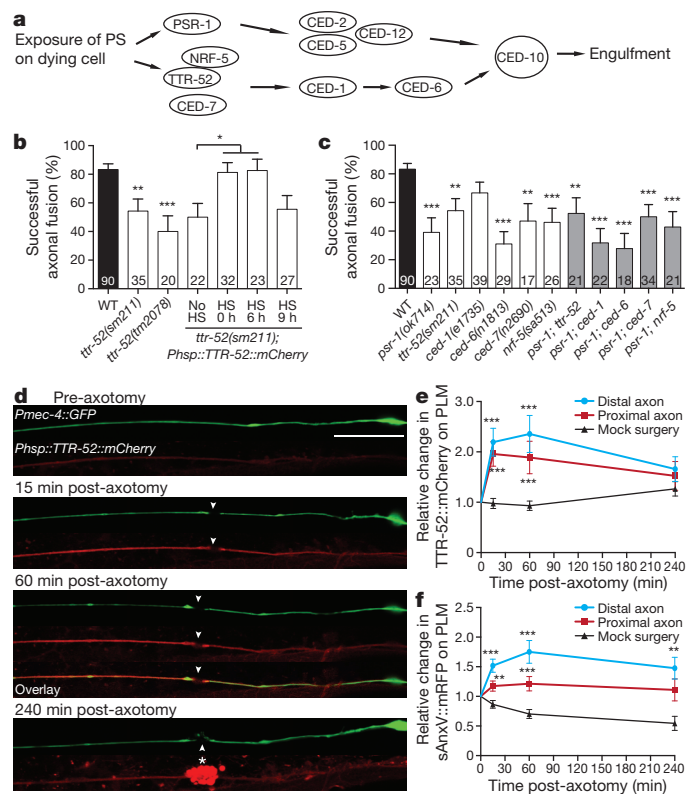


Figure 3 | PSR-1 functions in the TTR-52 pathway during axonal fusion. **a**, Schematic of two partially redundant parallel pathways that mediate recognition and engulfment of apoptotic cells. **b**, TTR-52 is required for successful axonal fusion. Expression of TTR-52::mCherry rescues the defect when induced from a heat-shock (HS) promoter immediately after axotomy (0 h) and up to 6 h post-axotomy. **c**, Mutations in genes acting in the *ttr-52* pathway reduce axonal fusion, and double mutants with *psr-1* are not more defective than single mutants. For **b** and **c**, error bars indicate standard error of proportion. *n* values within each bar; *P* values from *t*-test: **P* < 0.05, ***P* < 0.01, ****P* < 0.001. **d**, Single plane confocal images showing the localization of TTR-52::mCherry in *ttr-52(sm211)* mutants to the PLM axon before axotomy and at several time points after transection. Filled arrowheads point to cut sites, asterisk highlights accumulation of TTR-52::mCherry at the injury site. Scale bar, 25 μ m. Image representative of 22 animals. **e**, **f**, Quantification of the relative fluorescence levels of TTR-52::mCherry (**e**) and sAnxV::mRFP (**f**) on the distal and proximal PLM axonal segments after transection compared to mock surgery conditions. Error bars indicate standard error. *n* values for distal and proximal *n* = 22 (**e**) and *n* = 26 (**f**), for mock surgery *n* = 23 (**e**) and *n* = 21 (**f**). *P* values from Kolmogorov–Smirnov test: ***P* < 0.01, ****P* < 0.001 compared to mock surgery.

Similar to the localization pattern of PSR-1, we observed that a functional tagged version of CED-6 (mRFP::CED-6) predominantly localized to mitochondria in PLM, and accumulated in the regrowing proximal segment after axotomy (Extended Data Fig. 7a–c). Next, we analysed localization of a tagged version of the ABC transporter, CED-7 (CED-7::mRFP), which could cell-autonomously rescue the axonal fusion defect (Extended Data Fig. 7a), and found a diffuse localization pattern throughout the PLM axon before, and at several time points after axotomy (Extended Data Fig. 7d). Finally, we analysed the localization of NRF-5, a lipid-binding molecule that is normally secreted from the body wall muscle cells during apoptotic clearance¹². Using a mCherry-tagged version of NRF-5 expressed from a heat-shock promoter, we found that NRF-5 binds to the PLM axon after axotomy (Extended Data Fig. 7e, f). This suggests that, similarly to TTR-52, NRF-5 may recognize and bind to the exposed PS signal following injury.

We have previously shown that, although axonal fusion is a highly specific process, non-specific fusion can occur when a second axon in close proximity to PLM is also severed⁷. The PLN axon fasciculates with

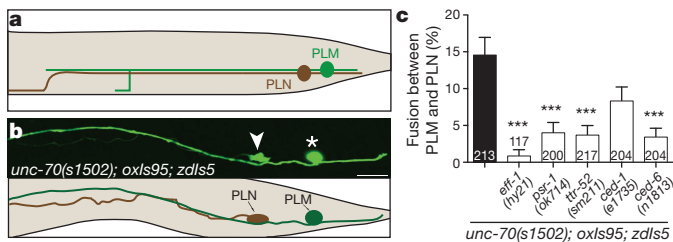


Figure 4 | Apoptotic molecules mediate non-specific axonal fusion caused by mutation of *unc-70*. **a**, Schematic showing fasciculation between the axons of PLN (brown) and PLM (green). **b**, Loss of *unc-70* causes spontaneous fusion between PLM (asterisk) and PLN (arrowhead) (appearance of GFP in PLN); single plane confocal image is shown. Scale bar, 10 μ m. Image representative of 31 animals. **c**, The frequency of non-specific fusion events is strongly reduced in animals with mutations in *eff-1*, *psr-1*, *ttr-52* and *ced-6*. Error bars indicate standard error of proportion. *n* values shown adjacent to each bar. *P* values from *t*-test: ****P* < 0.001 compared to *unc-70(s1502)*.

that of PLM (Fig. 4a), and when both axons are simultaneously transected approximately 10% of animals undergo non-specific fusion between these two neurons⁷. We observed a similar phenotype in animals deficient in UNC-70/ β -spectrin, a genetic background in which the axons of *C. elegans* motor neurons undergo spontaneous cycles of breaks followed by regeneration²⁹. We observed that *unc-70* mutant animals display aberrant PLM axonal morphology comprising breaks, loop structures, and branching (Extended Data Fig. 8a, b), as well as transfer of fluorophores between PLM and PLN (Fig. 4b, c). Time-lapse imaging and a tethered-fluorophore assay demonstrated that this transfer of fluorophores occurred through fusion events, and was not due to aberrant promoter activity (Extended Data Fig. 8c, d). These fusion events were dependent on the EFF-1 fusogen and required DLK-1, a mitogen-activated triple kinase essential for regeneration³⁰ (Fig. 4c and Extended Data Fig. 8e). Remarkably, we found that loss of *psr-1*, *ttr-52*, or *ced-6* caused a strong reduction in *unc-70*-induced PLM–PLN fusion (Fig. 4c). Analogous to our finding post-axotomy, we observed TTR-52::mCherry and sAnxV::mRFP binding to the PLM axon in *unc-70* mutants (Extended Data Fig. 9), suggesting that axonal damage induced by this genetic insult could also trigger the flipping of PS to the external surface of the membrane.

Thus, recognition of separated distal segments by a regrowing axon is achieved by changes in lipid membrane composition, and is mediated by apoptotic clearance molecules with the fusogen EFF-1 acting downstream to restore membrane and cytoplasmic continuity (Extended Data Fig. 10 and Supplementary Discussion).

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 13 December 2013; accepted 19 November 2014.

- Bedi, S. S. & Glanzman, D. L. Axonal rejoining inhibits injury-induced long-term changes in *Aplysia* sensory neurons *in vitro*. *J. Neurosci.* **21**, 9667–9677 (2001).
- Birse, S. C. & Bittner, G. D. Regeneration of giant axons in earthworms. *Brain Res.* **113**, 575–581 (1976).
- Deriemer, S. A., Elliott, E. J., Macagno, E. R. & Muller, K. J. Morphological evidence that regenerating axons can fuse with severed axon segments. *Brain Res.* **272**, 157–161 (1983).
- Ghosh-Roy, A., Wu, Z., Goncharov, A., Jin, Y. & Chisholm, A. D. Calcium and cyclic AMP promote axonal regeneration in *Caenorhabditis elegans* and require DLK-1 kinase. *J. Neurosci.* **30**, 3175–3183 (2010).
- Hoy, R. R., Bittner, G. D. & Kennedy, D. Regeneration in crustacean motoneurons: evidence for axonal fusion. *Science* **156**, 251–252 (1967).
- Macagno, E. R., Muller, K. J. & DeRiemer, S. A. Regeneration of axons and synaptic connections by touch sensory neurons in the leech central nervous system. *J. Neurosci.* **5**, 2510–2521 (1985).

- Neumann, B., Nguyen, K. C., Hall, D. H., Ben-Yakar, A. & Hilliard, M. A. Axonal regeneration proceeds through specific axonal fusion in transected *C. elegans* neurons. *Dev. Dyn.* **240**, 1365–1372 (2011).
- Wang, X. *et al.* Cell corpse engulfment mediated by *C. elegans* phosphatidylserine receptor through CED-5 and CED-12. *Science* **302**, 1563–1566 (2003).
- Liu, Q. A. & Hengartner, M. O. Candidate adaptor protein CED-6 promotes the engulfment of apoptotic cells in *C. elegans*. *Cell* **93**, 961–972 (1998).
- Wang, X. *et al.* *Caenorhabditis elegans* transthyretin-like protein TTR-52 mediates recognition of apoptotic cells by the CED-1 phagocyte receptor. *Nature Cell Biol.* **12**, 655–664 (2010).
- Wu, Y. C. & Horvitz, H. R. The *C. elegans* cell corpse engulfment gene *ced-7* encodes a protein similar to ABC transporters. *Cell* **93**, 951–960 (1998).
- Zhang, Y., Wang, H., Kage-Nakadai, E., Mitani, S. & Wang, X. *C. elegans* secreted lipid-binding protein NRF-5 mediates PS appearance on phagocytes for cell corpse engulfment. *Curr. Biol.* **22**, 1276–1284 (2012).
- Pérez-Vargas, J. *et al.* Structural basis of eukaryotic cell–cell fusion. *Cell* **157**, 407–419 (2014).
- Zeev-Ben-Mordehai, T., Vasishthan, D., Siebert, C. A. & Grunewald, K. The full-length cell–cell fusogen EFF-1 is monomeric and upright on the membrane. *Nat. Commun.* **5**, 3912 (2014).
- Mohler, W. A. *et al.* The type I membrane protein EFF-1 is essential for developmental cell fusion. *Dev. Cell* **2**, 355–362 (2002).
- Podbilewicz, B. Cell fusion. *WormBook* <http://dx.doi.org/10.1895/wormbook.1.52.1> (2006).
- Oren-Suissa, M., Hall, D. H., Treinin, M., Shemer, G. & Podbilewicz, B. The fusogen EFF-1 controls sculpting of mechanosensory dendrites. *Science* **328**, 1285–1288 (2010).
- Fadok, V. A. *et al.* Exposure of phosphatidylserine on the surface of apoptotic lymphocytes triggers specific recognition and removal by macrophages. *J. Immunol.* **148**, 2207–2216 (1992).
- Fadok, V. A. *et al.* A receptor for phosphatidylserine-specific clearance of apoptotic cells. *Nature* **405**, 85–90 (2000).
- Yang, H. *et al.* A lysine-rich motif in the phosphatidylserine receptor PSR-1 mediates recognition and removal of apoptotic cells. *Nat. Commun.* <http://dx.doi.org/10.1038/ncomms6717> (in the press).
- Chang, B., Chen, Y., Zhao, Y. & Bruick, R. K. JMJD6 is a histone arginine demethylase. *Science* **318**, 444–447 (2007).
- Hong, X. *et al.* Interaction of JMJD6 with single-stranded RNA. *Proc. Natl Acad. Sci. USA* **107**, 14568–14572 (2010).
- Webby, C. J. *et al.* Jmjd6 catalyses lysyl-hydroxylation of U2AF65, a protein associated with RNA splicing. *Science* **325**, 90–93 (2009).
- Gumienny, T. L. *et al.* CED-12/ELMO, a novel member of the CrkII/Dock180/Rac pathway, is required for phagocytosis and cell migration. *Cell* **107**, 27–41 (2001).
- Reddien, P. W. & Horvitz, H. R. CED-2/CrkII and CED-10/Rac control phagocytosis and cell migration in *Caenorhabditis elegans*. *Nature Cell Biol.* **2**, 131–136 (2000).
- Wu, Y. C. & Horvitz, H. R. *C. elegans* phagocytosis and cell-migration protein CED-5 is similar to human DOCK180. *Nature* **392**, 501–504 (1998).
- Zhou, Z., Hartwig, E. & Horvitz, H. R. CED-1 is a transmembrane receptor that mediates cell corpse engulfment in *C. elegans*. *Cell* **104**, 43–56 (2001).
- Mapes, J. *et al.* CED-1, CED-7, and TTR-52 regulate surface phosphatidylserine expression on apoptotic and phagocytic cells. *Curr. Biol.* **22**, 1267–1275 (2012).
- Hammarlund, M., Jorgensen, E. M. & Bastiani, M. J. Axons break in animals lacking β -spectrin. *J. Cell Biol.* **176**, 269–275 (2007).
- Hammarlund, M., Nix, P., Hauth, L., Jorgensen, E. M. & Bastiani, M. Axon regeneration requires a conserved MAP kinase pathway. *Science* **323**, 802–806 (2009).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank L. Hammond and S. Roy for support with microscopy; R. Tweedale for comments on the manuscript; P. Bazzicalupo, D. Hall, K. Nguyen, and members of the Hilliard laboratory for discussions and comments; and E. Jorgensen, S. Shaham and X. Wang for sharing reagents. Some strains were provided by the CGC, which is funded by NIH Office of Research Infrastructure Programs (P40 OD010440), and the International *C. elegans* Gene Knockout Consortium. This work was supported by NHMRC Project Grants 631634 and 1068871, NIH R01 NS060129, and ARC Future Fellowship FT110100097 to M.A.H.; NIH grants GM059083, GM079097 and GM088241 to D.X.; HFSPO Fellowship LT000762/2012 to R.G.S.; Australian Postgraduate Award to S.C.; The University of Queensland Research Scholarship to C.L.; and ARC LIEF grant LE130100078 for microscopy.

Author Contributions B.N., S.C., and R.G.S. carried out most experiments. C.L., E.S.L., and A.N. contributed some experiments. D.X. provided reagents, designed PSR-1 experiments, and edited the paper. B.N. and M.A.H. designed and interpreted experiments and wrote the paper.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to M.A.H. (m.hilliard@uq.edu.au).

METHODS

Strains and genetics. Standard techniques were used for *C. elegans* maintenance, crosses and other genetic manipulations³¹. All experiments were performed on hermaphrodites grown at room temperature ($\sim 22^{\circ}\text{C}$) unless otherwise stated, and all strains were grown on OP50 bacteria, except for animals carrying the *unc-70(s1502)* mutation, which were grown on HB101 bacteria. The following mutations were used: LGI: *ced-1(e1735)*, *ced-12(bz187)*, *dlk-1(jw476)*, *sem-4(n1971)*; LGII: *eff-1(ok1021)*, *eff-1(hy21)*; LGIII: *ced-4(n1162)*, *ced-6(n1813)*, *ced-7(n2690)*, *ced-9(n1950)*, *ttr-52(sm211)*, *ttr-52(tm2078)*; LGIV: *ced-2(e1752)*, *ced-3(n717)*, *ced-5(n1812)*, *ced-10(n3246)*, *psr-1(ok714)*, *psr-1(tm469)*; LGV: *nrf-5(sa513)*, *unc-70(s1502)*. The transgenes used were: *smls119[Phsp16-2::ttr-52::mCherry]¹⁰*, *smls95[Phsp16-2::sAnxV::mRFP]³²*, *oxIs95[Ppdi-2::unc-70; Pmyo2::GFP]³³*, *qxIs92[Phsp::nrf-5::mCherry]* (a gift from Xiaochen Wang), *vdEx575[Pmec-4::psr-1(c) (10 ng μl^{-1}), Podr-1::DsRed (30 ng μl^{-1}), vdEx661/vdEx662/vdEx663[Pmec-4::EFF-1::GFP (5 ng μl^{-1}); Pmec-4::mCherry (20 ng μl^{-1}); Podr-1::DsRed (60 ng μl^{-1}), vdEx705[Pmec-4::psr-1(c) (H192A/D194A) (10 ng μl^{-1}); Podr-1::DsRed (30 ng μl^{-1}), vdEx709[Pmec-4::psr-1(c)(K308E/K315E) (10 ng μl^{-1}); Podr-1::DsRed (30 ng μl^{-1}), vdEx851[Pmec-4::eff-1 (5 ng μl^{-1}); Podr-1::DsRed (60 ng μl^{-1}), vdEx862[Pmec-4::psr-1(c)::mCherry (2.5 ng μl^{-1}); Podr-1::DsRed (25 ng μl^{-1}), vdEx876[Pmec-4::tomm-20::GFP (0.5 ng μl^{-1}); Podr-1::GFP (50 ng μl^{-1}), vdEx895[Pmec-4::mRFP::ced-6 (5 ng μl^{-1}); Podr-1::DsRed (30 ng μl^{-1}), vdEx998[Pmec-4::ced-7::mRFP (2.5 ng μl^{-1}); Podr-1::DsRed (30 ng μl^{-1}), zdlIs5[Pmec-4::GFP].*

Molecular biology. Standard molecular biology techniques were used. The *Pmec-4::EFF-1* plasmid was created by excising *EFF-1* from a *Podr-1::EFF-1* plasmid using Kpn I and Apa I restriction enzymes, and cloning it in front of the *Pmec-4* promoter in a *Pmec-4::unc-54_3'* UTR plasmid. To build the *Pmec-4::EFF-1::GFP* plasmid, *eff-1* genomic DNA was amplified (fwd primer 5'-ctagtagcatggaacccgcttgtagtg-3'; rev primer 5'-gaaccgggttaagtactgctactgctatag-3') and cloned into the pSM::mCherry plasmid using Nhe I and Age I restriction enzymes, to obtain pSM::*EFF-1::mCherry*. The *mec-4* promoter was then cloned from plasmid pSM::*Pmec-4::EFF-1* into pSM::*EFF-1::mCherry* using Sal I and Sph I restriction enzymes, to obtain *Pmec-4::EFF-1::mCherry*. Finally, an Age I/Apa I fragment from plasmid pCP.179 (ref. 33) (a gift from Shai Shaham) was cloned into *Pmec-4::EFF-1::mCherry* to obtain *Pmec-4::EFF-1::GFP*. The *Pmec-4::psr-1(c)* plasmid was generated through insertion of a Bam HI/Age I *psr-1* cDNA amplicon between the *mec-4* promoter and the *unc-54_3'* UTR. The following primers were used: fwd primer 5'-tcagtgggcatcatgctattggcgagatag-3'; rev primer 5'-tcagtggcgggttaaac tagcaggttgtaaaat-3'. The *Pmec-4::psr-1(c)::mCherry* plasmid was created in similar fashion, except a different reverse primer was used (rev primer 5'-tcagtggcgggtc gaactagcaggttgtaaaatag-3') and the cDNA was cloned into a *Pmec-4::mCherry::unc-54_3'* UTR plasmid. QuikChange mutagenesis (Agilent Technologies, Santa Clara, CA) was used to introduce the K308/K315E and H192A/D194A mutations, with the *Pmec-4::psr-1(c)* plasmid used as the template. The *ced-6* gene was amplified from genomic DNA and inserted downstream of mRFP into a *Pmec-4::mRFP::unc-54_3'* UTR plasmid using Xma I and Age I restriction enzymes. The following primers were used: fwd primer 5'-tcagtcgccggaatggcaaaagacatttataca gacc-3'; rev primer 5'-tcagtcacgggtttattgctcaaaattcatctttcg-3'. The *Pmec-4::tomm-20::GFP* plasmid was created by switching the mRFP fluorophore from a *Pmec-4::tomm-20::mRFP³⁴* plasmid using Age I and Eco RI restriction enzymes. *Pmec-4::ced-7::mRFP* was created through insertion of full-length *ced-7* cDNA upstream of the mRFP sequence in a *Pmec-4::mRFP::unc-54_3'* UTR plasmid using Xma I and Nhe I restriction enzymes and the following primers: fwd primer 5'-tcagtcgccg gataatagattggcacaattctc-3', and rev primer 5'-tcagtgctagcgacatgtgtggaatgggaaa tc-3'.

Laser axotomy and microscopy. Animals were immobilized in 0.05% tetramisole hydrochloride on 4% agar pads. Axotomies were performed as previously described⁷, using a MicroPoint Laser System Basic Unit attached to a Zeiss Axio Imager A1 (Objective EC Plan-Neofluar 100 \times /1.30 Oil M27). Axons were severed in L4 or one-day-old adult (for localization of TTR-52, sAnxV, and NRF-5) animals approximately 50 μm anterior to the PLM cell body, visualized on a Zeiss Axio Imager Z1 microscope equipped with a Photometrics camera (Cool Snap HQ2; Tucson, AZ) and analysed with MetaMorph software (Molecular Devices, Sunnyvale, CA) or ImageJ 1.48s. Animals were analysed 24 h post-axotomy for regrowth, reconnection and fusion. Regrowth was recorded as the length of regenerative growth from the proximal side of the cut site: the proximal axon was traced from the start of regrowth to the tip of the longest regenerative branch using MetaMorph software; neurons that underwent axonal fusion were excluded from these quantifications. Axons were deemed to be reconnected when the proximal and distal axons were visually connected and within the same focal plane when observed with a 63 \times objective (Extended Data Fig. 1). Axonal fusion was defined as proximal–distal reconnection that prevented the onset of degeneration in the distal axon⁷. For clarification of successful fusion, animals were re-analysed at 48 h and/or 72 h post-axotomy as necessary.

To analyse a selection of reconnection events with greater resolution, three-dimensional structured illumination microscopy (3D-SIM) was performed with an Elyra PS.1 SIM/STORM microscope, using a 63 \times 1.4 NA oil objective. To visualize reconnection sites, z-stacks (3D-SIM) were acquired and SIM processing performed. Green fluorescence was visualized with a 488 nm laser (10% power, 50 ms exposure time, 28 μm grating, 3 rotations). Processing was completed with a sectioning of 100/83/83. Wild-type and *eff-1(ok1021)* mutant animals were analysed 24 h post-axotomy and scored for proximal–distal reconnection with both standard compound imaging and 3D-SIM. Reconnection was scored separately and double blinded with respect to each technique before comparisons were made to determine the accuracy between the two techniques. In cases where the distal fragment had become too faint, SIM processing could not be completed.

Visualization of EFF-1. To study the localization of EFF-1::GFP in regrowing axons, laser axotomies were performed in animals carrying the *vdEx662* transgene in 0.05% tetramisole. Acquisitions were performed on animals mounted on 10% agarose pads in M9, in 25 mM sodium azide. Individual animals were then imaged 30 min, 3 h, 8 h and 18–24 h after axotomy using a LSM 710 META confocal microscope, equipped with a GaAsP detector and Zen 2012 software. Green fluorescence was visualized with a 488 nm laser (10% power, with a gain of 600 and 4 \times averaging) and red fluorescence was visualized with a 543 nm laser (1.4% power, with a gain of 500 and 4 \times averaging). Image analysis was performed using ImageJ 1.48s. To score for the presence of EFF-1::GFP on the membrane of regrowing axons, fluorescence profiles of line scans at different sections of the regrowing axon tip were obtained using the 'Plot Profile' tool in ImageJ. EFF-1::GFP was scored as localizing to the membrane if the peak of green intensity (EFF-1::GFP) and the peak of red intensity (cytoplasmic mCherry) were not overlapping.

To visualize EFF-1::GFP by SIM, axotomized animals were mounted in 10% agarose pads in M9, in 25 mM sodium azide. Single plane acquisitions (2D-SIM) were performed with an Elyra PS.1 SIM/STORM microscope, using a 63 \times 1.4 NA oil objective, and SIM processing was performed using the ZEN software 2012, Black edition. Green fluorescence was visualized with a 488 nm laser (20% power, 200–400 ms exposure time, 28 μm grating, 3 rotations) and red fluorescence was visualized with a 561 nm laser (10% power, 200 ms exposure time, 34 μm grating, 3 rotations). Processing was completed with a sectioning of 100/83/83. Fluorescence profiles of line scans were obtained as described above.

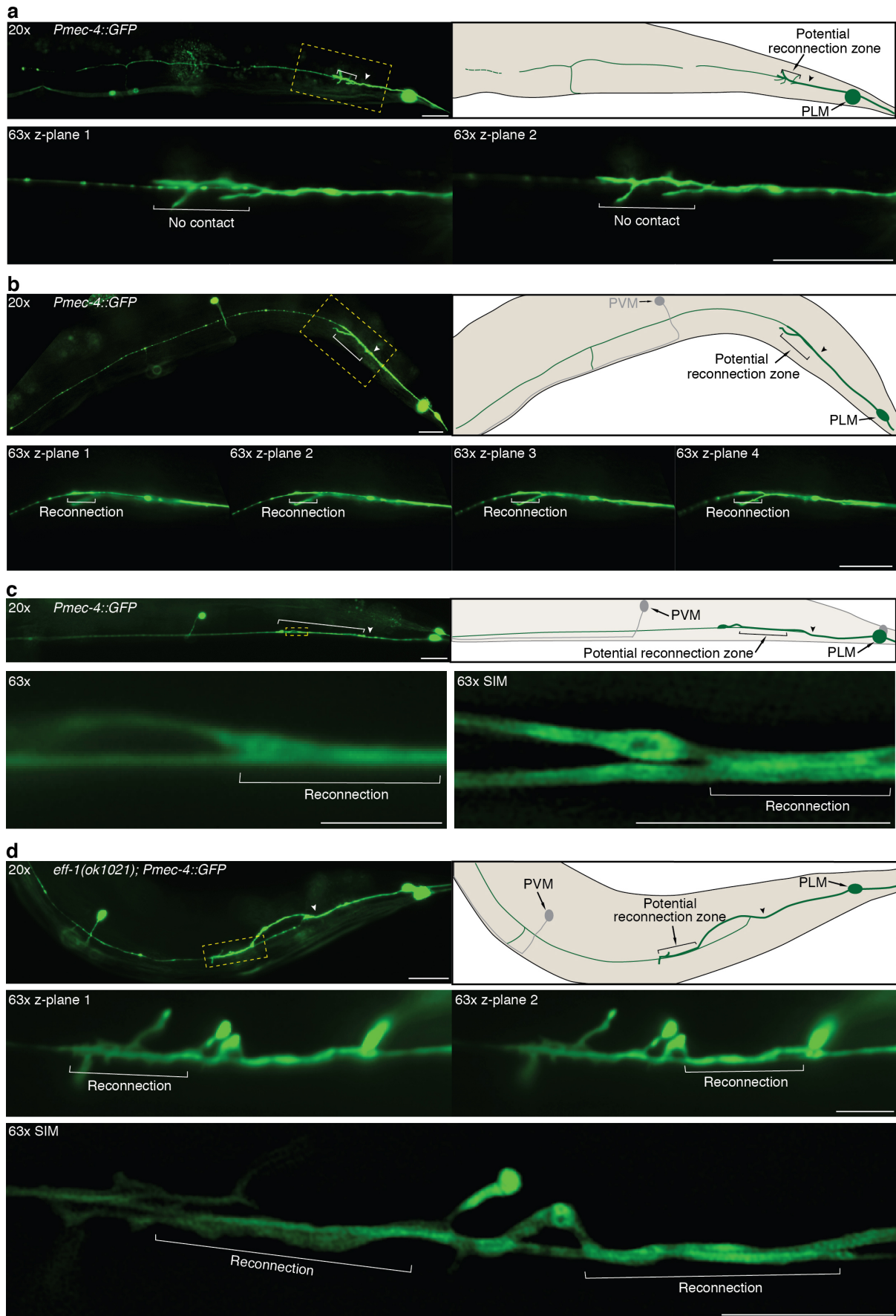
Visualization of other tagged proteins. The subcellular localization of PSR-1::mCherry, mRFP::CED-6, CED-7::mRFP, and NRF-5::mCherry was analysed in animals immobilized with 0.05% tetramisole with a LSM 710 META confocal microscope, equipped with a GaAsP detector and Zen 2012 software. Green cytoplasmic fluorescence was visualized with a 488 nm laser (0.15% power, with a gain of 650 and 4 \times averaging) and red fluorescence was visualized with a 561 nm laser (1.5% power, with a gain of 700 and 4 \times averaging for PSR-1 and CED-6; 0.7% power, with a gain of 600 and 4 \times averaging for CED-7; 2.4% power, with a gain of 700 and 8 \times averaging for NRF-5). TTR-52::mCherry and sAnxV::mRFP were visualized in one-day-old adult animals (unless stated otherwise) with a LSM 510 META confocal microscope and Zen 2008 software. Green fluorescence was analysed with a 488 nm laser (2.4% power, with a gain of 601 and 8 \times averaging) and red fluorescence with a 543 nm laser (100% power, with a gain of 1012 and 8 \times averaging).

Quantification of the binding of TTR-52, Annexin V, and NRF-5 to the PLM axon. Expression of transgenes was induced 4 h before analysis with incubation at 30 $^{\circ}\text{C}$ for 30 min. Changes in fluorescence intensity were calculated with line scans performed using ImageJ 1.46r software. To avoid bright fluorescence associated with collateral laser damage, line scans were recorded $\sim 5 \mu\text{m}$ from the damage site on both the proximal and distal axon segments for a length of $\sim 15 \mu\text{m}$. Line scans were recorded along the same region for the 488 nm and 543 nm channels for each image. To control for changes in focus and intensity, mCherry/mRFP expression was normalized to GFP for each image and recorded relative to pre-axotomy levels. For the *unc-70* mutant background, analyses were performed as above, except that for TTR-52::mCherry line scans were recorded along the PLM axon for a length of $\sim 15 \mu\text{m}$, whereas for sAnxV::mRFP three separate line scans were recorded along the PLM axon for a length of $\sim 5 \mu\text{m}$, with the mean values used for each axon. Background fluorescence of the 488 nm and 543 nm channels was also subtracted from the corresponding axonal line scans.

Analysis of cell–cell fusion events in *unc-70* mutants. For time-lapse microscopy one-day-old adult *unc-70(s1502)* animals were immobilized in 0.05% tetramisole hydrochloride on 10% agar pads sealed with Vaseline and visualized using a Zeiss inverted spinning disk confocal microscope equipped with a Yokogawa W1 disk head and a Piezo z-drive. Images were acquired every 30 min for 15 h with a Hamamatsu Flash 4.0 scientific CMOS camera using Slidebook 5.0 software. Adult animals were used for the tethered fluorophore experiments, and the presence and location of TOMM-20::mRFP were scored in the PLM neuron, and GFP-positive PLN and additional neurons. For *unc-70(s1502)* animals, only those displaying PLM–PLN fusion were analysed.

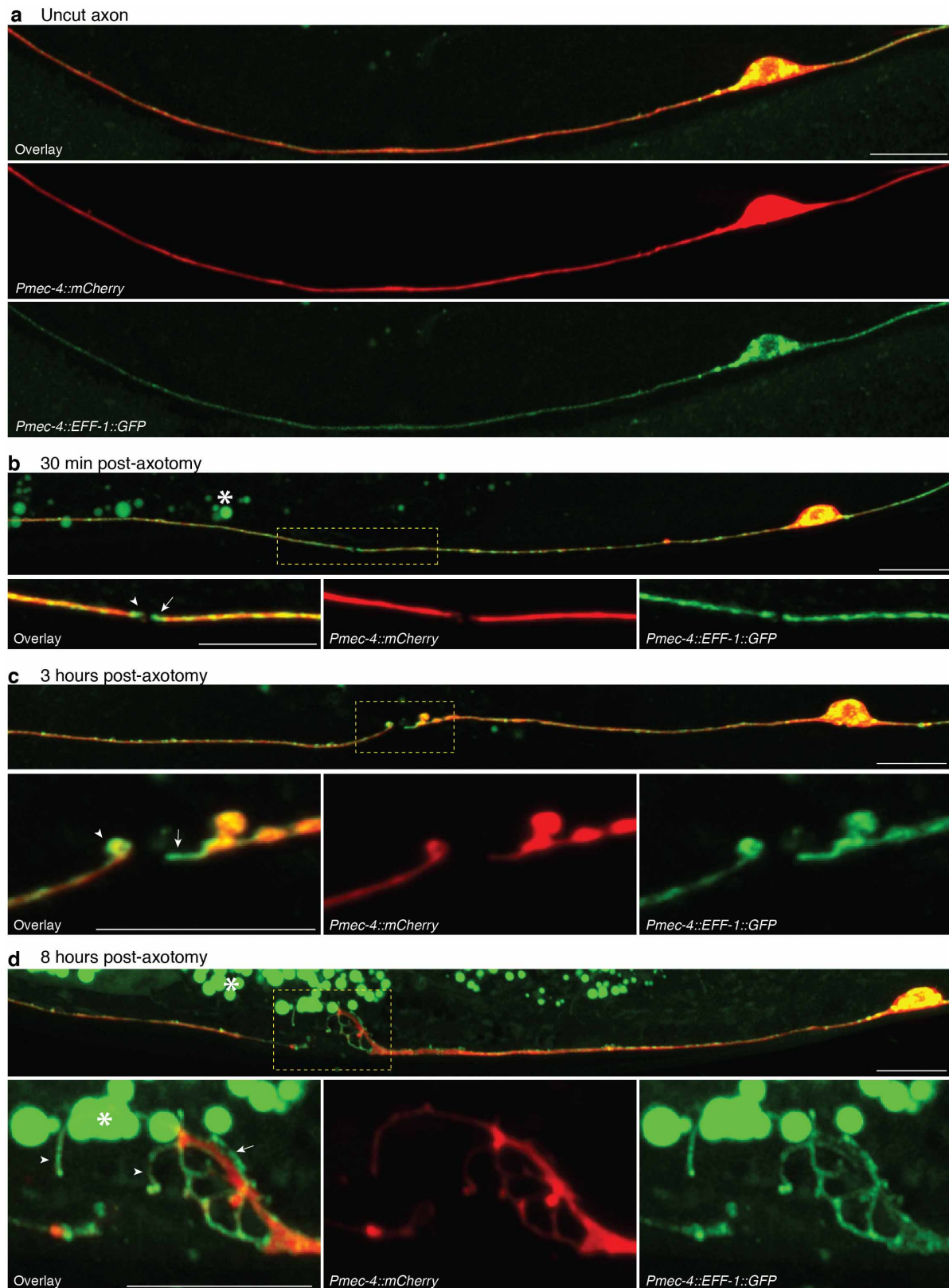
Statistical analysis. Statistical analysis was performed using Primer of Biostatistics 3.01, GraphPad Prism and Microsoft Excel. Error of proportions was used to assess variation across a single population, two-way comparison was performed using the *t*-test, and ANOVA was used for more than two groups. For those data sets that did not follow a Normal distribution we used the Kolmogorov–Smirnov test (Fig. 3e, f and Extended Data Figs 5d–f, 6a, 7f, and 9c, f). No statistical method was used to predetermine sample size.

31. Brenner, S. The genetics of *Caenorhabditis elegans*. *Genetics* **77**, 71–94 (1974).
32. Zou, W. *et al.* *Caenorhabditis elegans* myotubularin MTM-1 negatively regulates the engulfment of apoptotic cells. *PLoS Genet.* **5**, e1000679 (2009).
33. Procko, C., Lu, Y. & Shaham, S. Glia delimit shape changes of sensory neuron receptive endings in *C. elegans*. *Development* **138**, 1371–1381 (2011).
34. Neumann, B. & Hilliard, M. A. Loss of MEC-17 leads to microtubule instability and axonal degeneration. *Cell Rep.* **6**, 93–103 (2014).
35. Mitani, S., Du, H., Hall, D. H., Driscoll, M. & Chalfie, M. Combinatorial control of touch receptor neuron expression in *Caenorhabditis elegans*. *Development* **119**, 773–783 (1993).



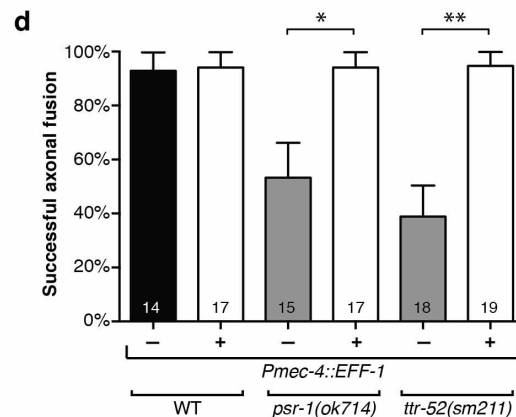
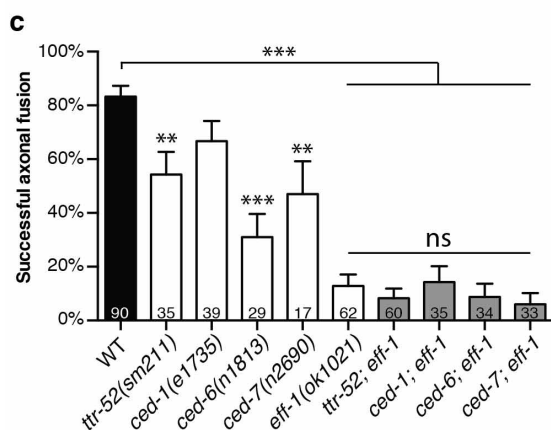
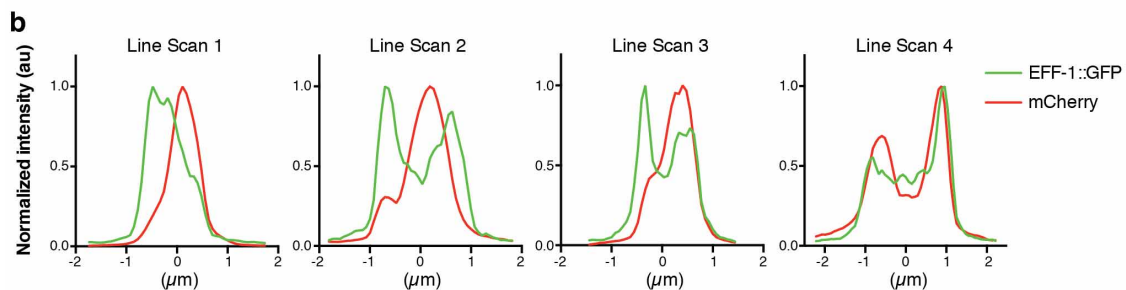
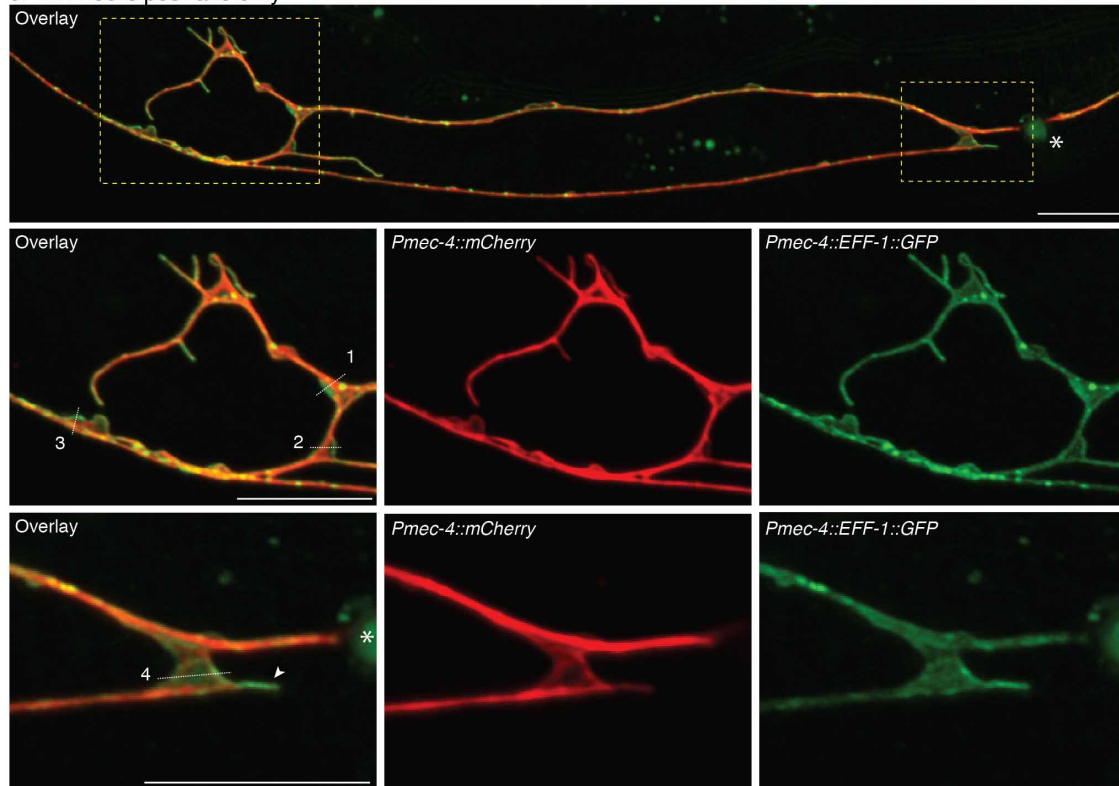
Extended Data Figure 1 | Analysis of proximal–distal reconnection using compound microscopy and 3D-SIM imaging. Typical examples of how reconnection was scored with compound microscopy (**a, b**). **a**, Image and schematic of a wild-type (WT) animal 24 h post-axotomy captured with a 20 \times objective (top panel) and two different *z* planes captured with a 63 \times objective. Despite visually appearing reconnected at 20 \times , the proximal regrowing axon has not made contact with the distal axon (63 \times); image representative of 150 animals. **b**, Representative images from a total of 90 animals taken with a 20 \times objective (top panels) and a 63 \times objective (bottom panels) of a WT animal 24 h post-axotomy in which reconnection has occurred between the proximal and distal axon segments. Three dimensional structured illumination microscopy (3D-SIM) was used to confirm contact regions between regrowing and distal axon fragments in WT (**c**) and *eff-1(ok1021)* mutant (**d**) animals. In each case that reconnection was observed with compound microscopy (63 \times), it was also evident with 3D-SIM imaging; *n* = 14 for each genotype. **c**, Successful reconnection in a WT animal, 24 h

post-axotomy. Top panel: single-plane image acquired with a 20 \times objective on a compound microscope. Bottom panels: single-plane images acquired with a 63 \times objective on a compound microscope (left, 63 \times) and a SIM image of the same region (right, 63 \times SIM); images representative of 18 animals. **d**, Proximal–distal reconnection in an *eff-1(ok1021)* mutant animal, 24 h post-axotomy. Top panel: single-plane image acquired with a 20 \times objective on a compound microscope. Middle panel: single-plane images acquired with a 63 \times objective on a compound microscope (63 \times), showing two different *z*-planes where the regions indicated by brackets as ‘contact’ zones are in focus. Bottom panel: a SIM image of the same regions highlighting the points of contact; images representative of 14 animals. Arrowheads point to the site of axotomy, and brackets highlight zones of either potential (**a**) or actual reconnection (**b–d**). Dashed boxes represent regions magnified in the 63 \times images. Scale bars, 25 μ m (**a, b** and 20 \times images in **c, d**) or 5 μ m (63 \times images in **c, d**). Estimated resolution of SIM images (as per Zeiss software ZEN Black 2012): 120 nm in (**c**) and 110 nm in (**d**).



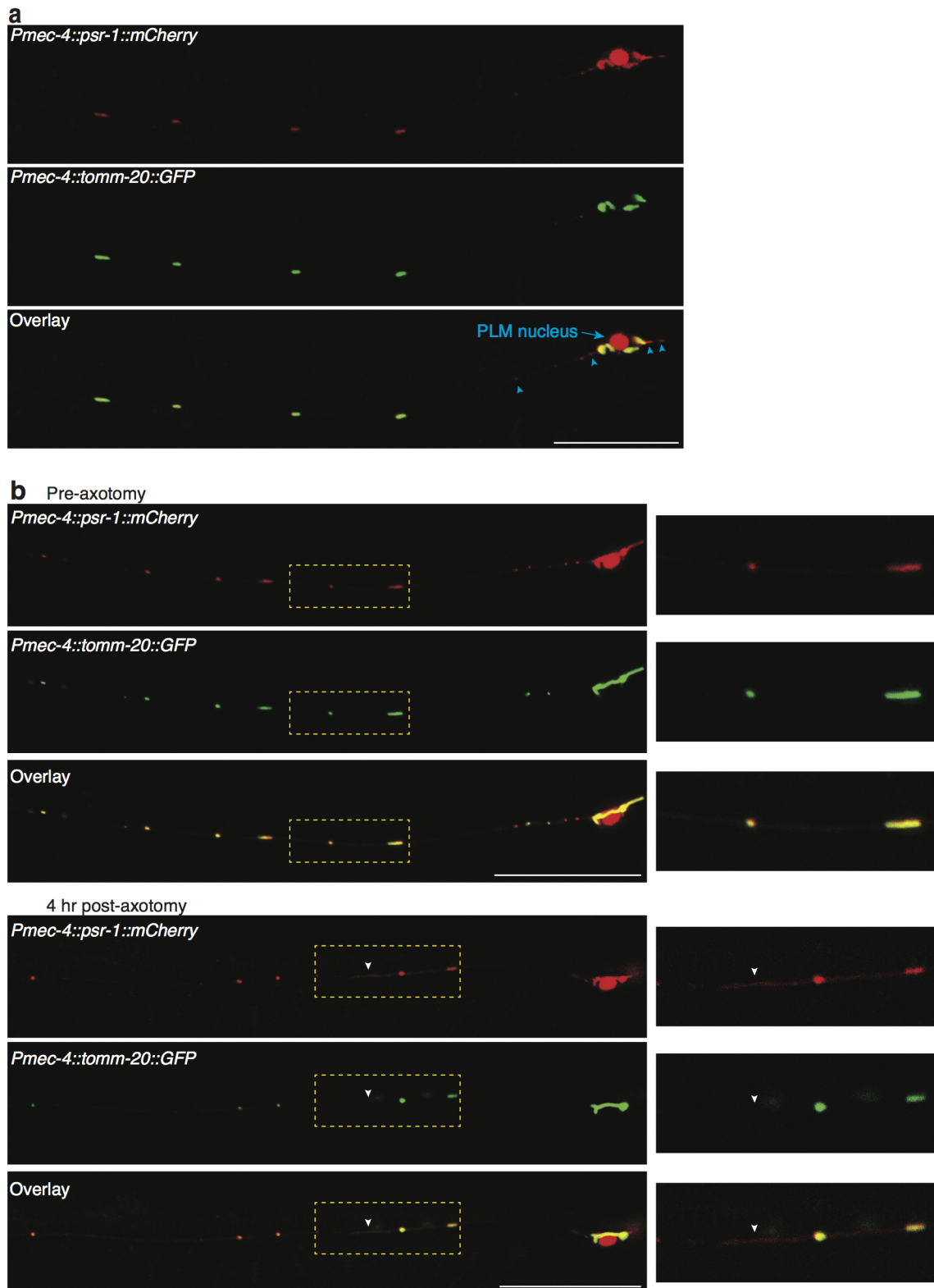
Extended Data Figure 2 | EFF-1 localization in PLM neurons before and after axotomy. Maximum projections of z-stack confocal images of an uncut axon and regrowing axons carrying the transgene *Pmec-4::EFF-1::GFP*; *Pmec-4::mCherry* in an *eff-1(ok1021)* mutant background. Representative images are shown for an intact axon (a), and axotomized axons 30 min post-axotomy (b), 3 h post-axotomy (c), and 8 h post-axotomy (d). a, In uncut axons, EFF-1 is present in the cell body (excluding the nucleus), posterior process and all along the axon in an irregular pattern. b, Thirty minutes after axotomy, EFF-1 forms a

'cap' on both the proximal (arrow) and distal (arrowhead) fragments. c, After 3 h, the proximal stump begins to swell. 'Capping' of the proximal (arrow) and distal (arrowhead) fragments by EFF-1 is still present. d, After 8 h, a growth cone with filopodia-like protrusions forms from the proximal stump. EFF-1 is present on the membrane of the growth cone (arrow) and filopodia-like structures (arrowheads). Scale bars, 10 μ m. Asterisks highlight intestinal autofluorescence. Images representative of 15 (a), 12 (b), 9 (c), and 7 (d) animals.

a 24 hours post-axotomy**Extended Data Figure 3 | EFF-1 localizes to the membrane of growth cones, and functions genetically downstream from the apoptotic genes.**

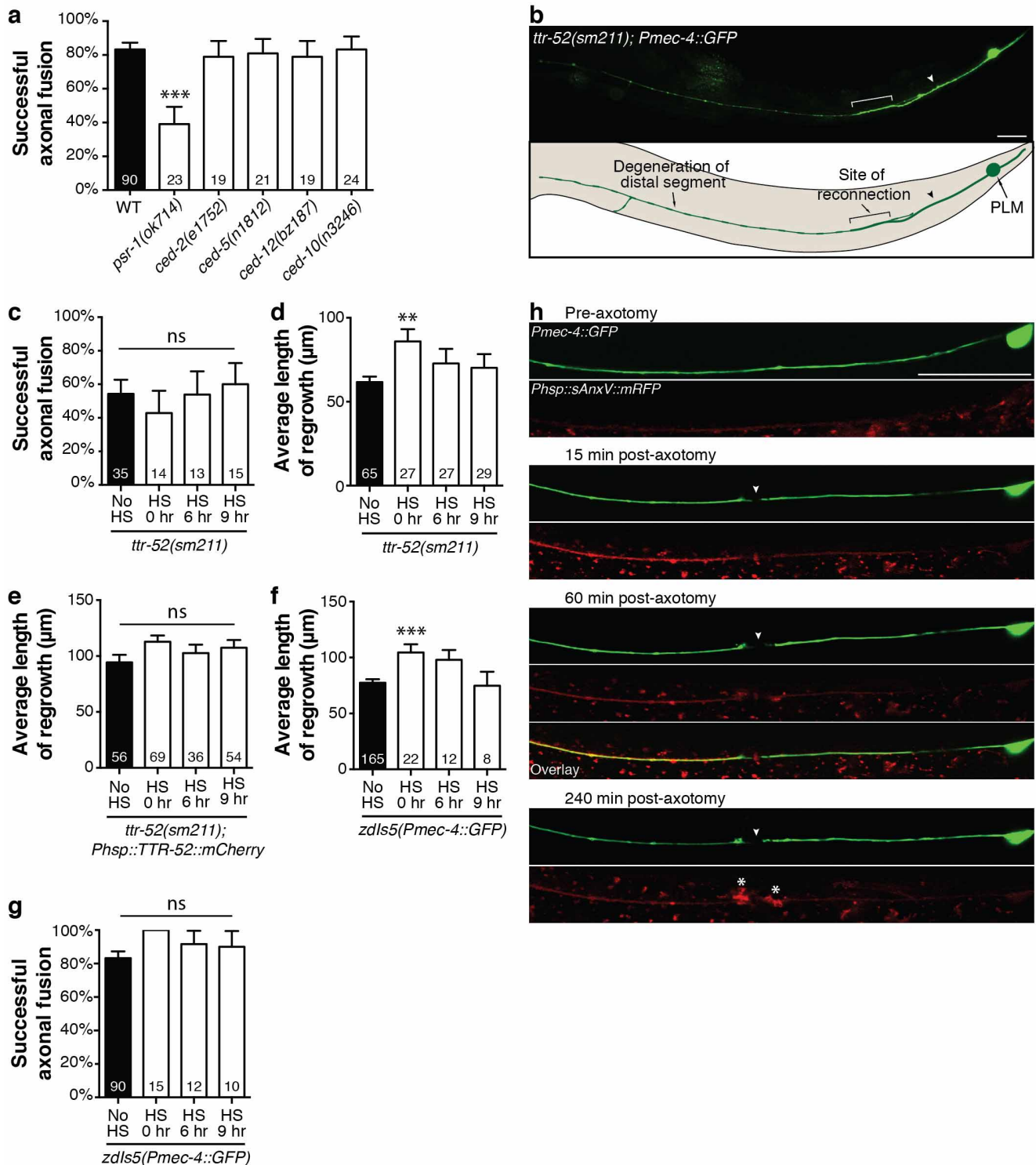
a, Maximum projection of z-stack confocal images of a regrowing axon carrying the transgene *Pmec-4::EFF-1::GFP*; *Pmec-4::mCherry* in *eff-1(ok1021)* mutant background, 24 h post-axotomy. EFF-1 is present on the membrane of the growth cone (line scans 1, 2 and 3) and forms a 'cap' at the proximal tip of the distal fragment (arrowhead). Asterisk shows the scar at the cut site. Scale bars, 10 μ m. Images representative of 8 animals. **b**, Fluorescence profiles from the line scans shown in the left panels of panel **a** illustrate the presence of EFF-1::GFP on the membrane of the growth cone (1 to 3) but not on the

membrane of a putative fusion site (line scan 4). au, arbitrary units. **c**, Successful axonal fusion in single mutant animals (white bars) compared to double mutants with *eff-1(ok1021)* (grey bars). Double mutants do not show a worsening of the defect compared to *eff-1* single mutants. **d**, Cell-autonomous overexpression of EFF-1 (white bars) in wild-type (WT) animals, or in those carrying mutations in *psr-1* or *ttr-52*, rescuing their defect. For **c** and **d**, error bars represent standard error of proportion. *n* values shown adjacent to each bar. *P* values from *t*-test: **P* < 0.05, ***P* < 0.01, ****P* < 0.001 compared to WT unless otherwise indicated. ns, not significant.



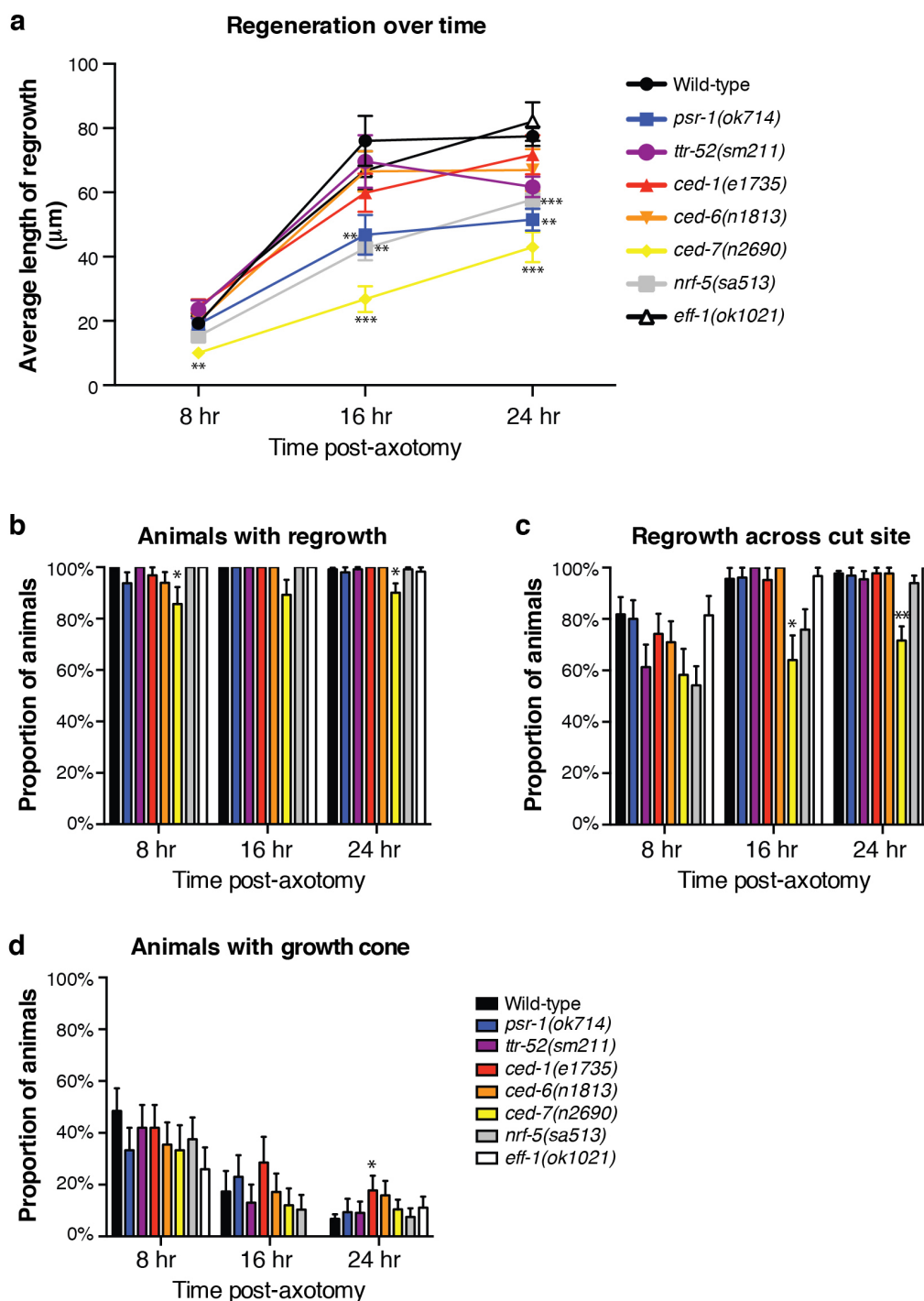
Extended Data Figure 4 | Analysis of the localization of functional fluorophore-tagged PSR-1. **a**, Maximum projection confocal images showing the co-localization of PSR-1::mCherry (top panel) with a mitochondrial marker (TOMM-20::GFP, middle panel) in a *psr-1(ok714)* mutant animal. PSR-1 displays strong co-localization with the mitochondria; however, there are regions where co-localization is not present, and these are highlighted by the arrow and arrowheads in the overlay image (bottom panel). Images are

representative of 35 animals. Animals of the same genotype were analysed before and after axotomy (**b**) to determine the localization of PSR-1 compared to the mitochondrial marker during regrowth. PSR-1 accumulates to the end of the regrowing proximal axon, whereas the TOMM-20::GFP remains restricted to the mitochondria; images representative of 27 animals. Panels to the right show magnified views of the boxed regions; arrowheads point to the site of axotomy. Scale bars, 25 μ m.



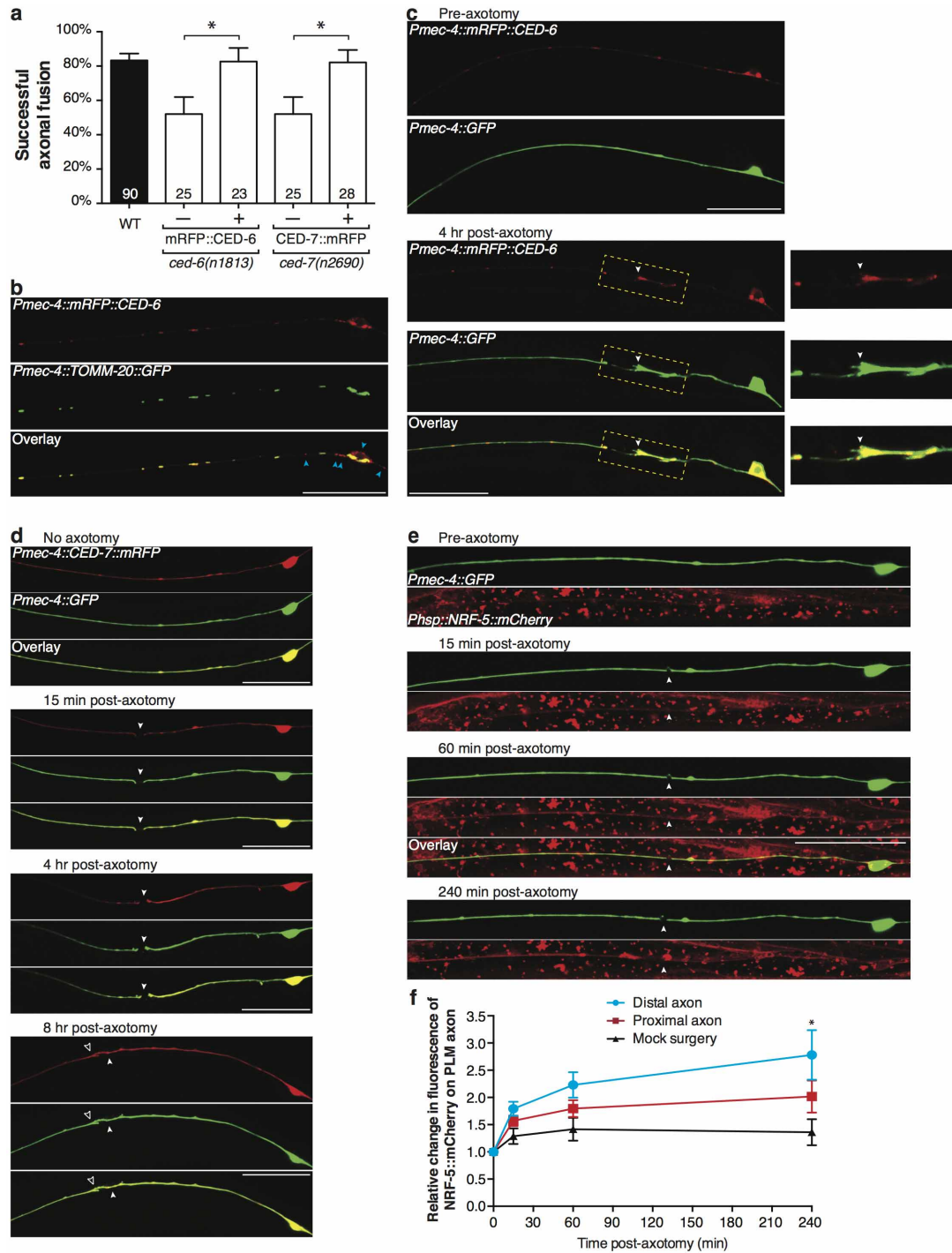
Extended Data Figure 5 | Analysis of axonal fusion and regeneration after heat-shock treatment in wild-type and *ttr-52* mutant animals, and analysis of sAnxV localization after axotomy. **a**, Animals mutant in *ced-2*, *ced-5*, *ced-12*, or *ced-10* have normal axonal fusion similar to that of wild-type (WT) animals. **b**, Unsuccessful fusion in a *ttr-52(sm211)* animal. Regrowth has extended from the cut site (arrowhead) and contacted the distal fragment (bracketed region), but degeneration of the distal fragment has begun. Scale bar, 25 μ m. Image representative of 16 animals. **c**, Rate of successful axonal fusion in *ttr-52(sm211)* animals without or with heat-shock (HS) treatment. **d**, The average length of regrowth in *ttr-52(sm211)* animals given HS after axotomy compared to those without HS treatment. **e**, Quantification of regrowth in *ttr-52(sm211); Phsp::TTR-52::mCherry* animals without HS treatment compared to those that received HS immediately after axotomy (0 h), or at 6 or

9 h post-axotomy. **f**, Comparison of the average length of regrowth in *zdl5* animals without HS or with HS given at different times after axotomy. **g**, Successful axonal fusion in *zdl5* animals without HS compared to animals given HS at different times post-axotomy. Error bars represent standard error of proportion (**a**, **c**, **g**) and standard error (**d**–**f**). *n* values shown adjacent to each bar. *P* values (from *t*-test for **a**, **c**, **g**; from Kolmogorov–Smirnov test for **d**–**f**): ***P* < 0.01, ****P* < 0.001 compared to WT or no HS (black bars), unless marked otherwise on the graph. ns, not significant. **h**, Single plane confocal images showing the localization of the PS sensor, sAnxV::mRFP, to the axon of PLM before axotomy and at several time points after transection. The overlay image shows the 60 min time-point; filled arrowhead points to the site of axotomy, and asterisks highlight the accumulation of sAnxV::mRFP at the injury site. Scale bar, 25 μ m. Images representative of 26 animals.



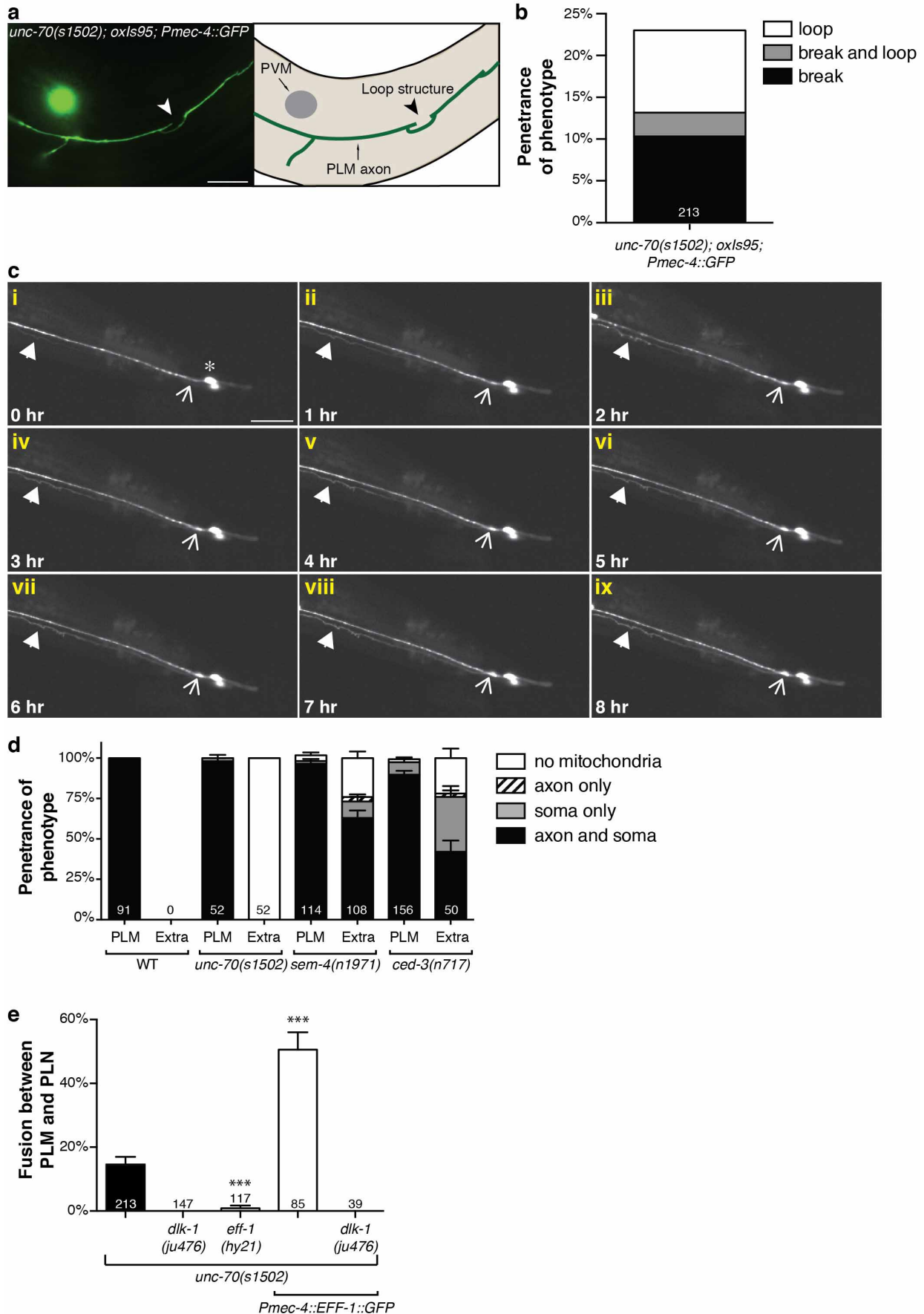
Extended Data Figure 6 | Quantification of regeneration. **a**, To determine if the loss of any of the genes analysed caused a reduction in the regenerative potential of PLM, we calculated the average length of regrowth over the first 24 h post-axotomy. Mutations in several genes that reduced axonal fusion, including *psr-1*, *ttr-52*, *nrf-5* and *ced-7*, caused defective regrowth in PLM. However, we found no correlation between the average length of regrowth and the rate of axonal fusion. For instance, *ced-10* and *ced-12* mutants displayed reduced regrowth (Supplementary Table 2), but presented no defect in axonal fusion (Extended Data Fig. 5a), whereas animals that lacked *eff-1* had the strongest defect in axonal fusion, but their regenerative growth was not

different from that of wild-type (WT) animals. Moreover, there was no correlation between axonal fusion and the percentage of animals able to initiate any regrowth from the proximal axon after transection (**b**), the average number of axons that were able to regrow across the cut site at different time points after axotomy (**c**), or the percentage of axons presenting a growth cone at different stages of regeneration (**d**). Error bars represent standard error (**a**) or standard error of proportion (**b–d**). *n* values are displayed in Supplementary Table 2. *P* values (from Kolmogorov–Smirnov test for **a**; from *t*-test for **b–d**): **P* < 0.05, ***P* < 0.01, ****P* < 0.001.



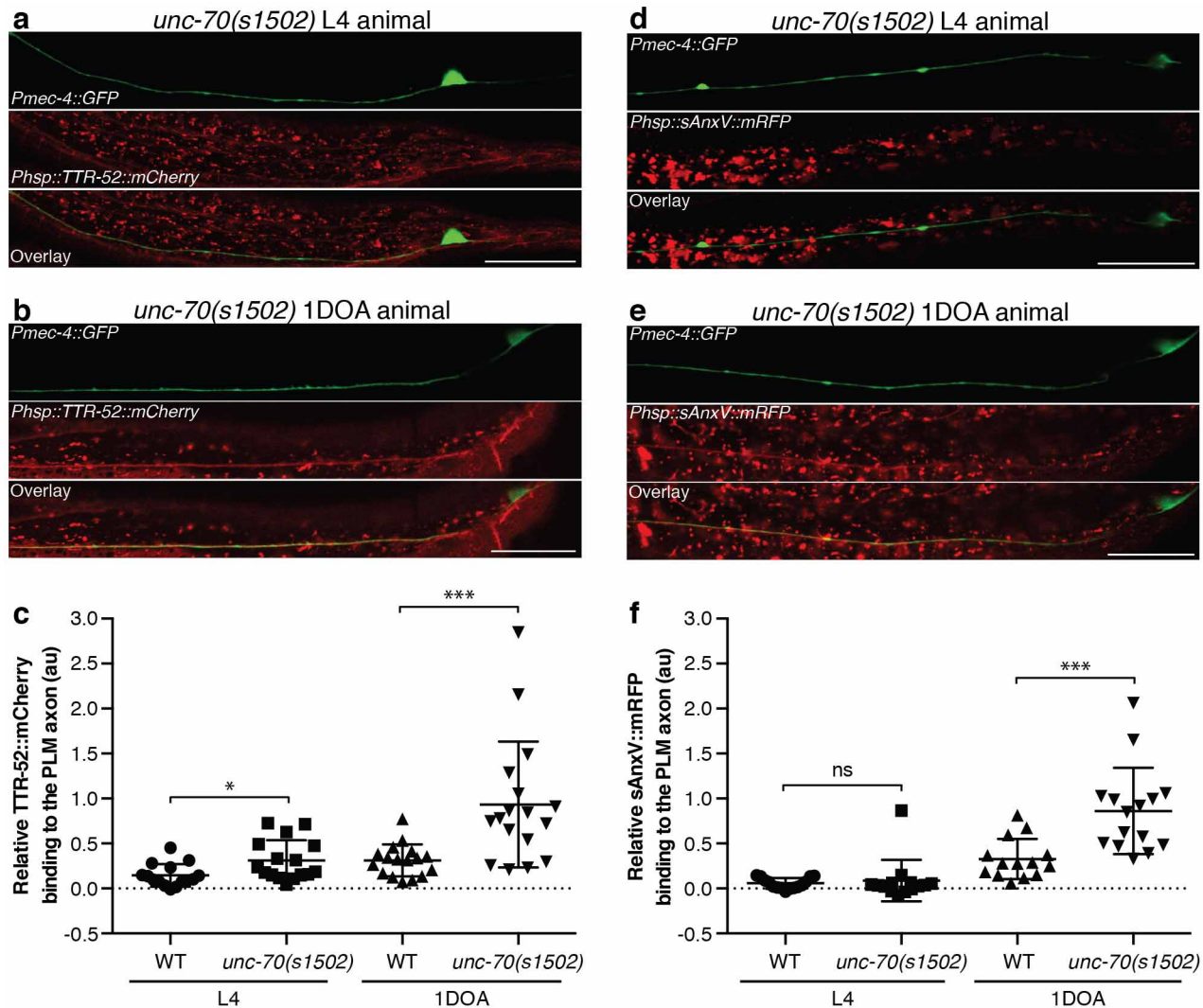
Extended Data Figure 7 | Localization of CED-6, CED-7, and NRF-5 in PLM before and after axotomy. **a**, Cell-autonomous rescue of the axonal fusion defect in *ced-6(n1813)* and *ced-7(n2690)* animals with expression in the mechanosensory neurons of either wild-type *ced-6* genomic DNA, or *ced-7* cDNA, tagged with mRFP. Error bars represent standard error of proportion; n values displayed in graph. P values from t -test: $*P < 0.05$. **b**, Maximum projection confocal images highlighting co-localization between mRFP::CED-6 and a mitochondrial marker (TOMM-20::GFP) in a *ced-6(n1813)* mutant background. Arrowheads point to regions where this co-localization is absent. Scale bar, 25 μ m. Images representative of 12 animals. **c**, Single plane confocal images showing the localization of mRFP::CED-6 in PLM in a *ced-6(n1813)* mutant animal before axotomy (top two panels), and 4 h post-axotomy (bottom three panels). Panels to the right show magnified images of the boxed regions to highlight accumulation of CED-6 at the end of the regrowing proximal axon. Arrowheads point to the site of axotomy. Scale bars, 25 μ m; images representative of 9 animals. **d**, CED-7::mRFP localization in single plane

confocal images of PLM in *ced-7(n2690)* mutant animals without axotomy (top panels) and at several time points after axotomy. Different animals are shown for each time point. Filled arrowheads point to the site of axotomy; open arrowhead points to site of fusion. Scale bars, 25 μ m. Images representative of 10 animals. **e**, Single plane confocal images of the localization of NRF-5::mCherry (expressed from a heat-shock promoter) in *nrf-5* mutants to the PLM axon before axotomy and at several time points after transection in animals with genotype *nrf-5(sa513); qxIs92; zdIs5*. Overlay image is shown for the 60 min time-point; arrowheads point to the site of axotomy. Scale bar, 25 μ m. Images representative of 29 animals. **f**, Quantification of the relative fluorescence levels of NRF-5::mCherry on the PLM distal (blue line) and proximal (red line) axons after axotomy compared to mock surgery conditions (black line). Error bars represent standard error; $n = 29$ for distal and proximal axon, $n = 13$ for mock surgery. P values from Kolmogorov–Smirnov test: $*P < 0.05$ compared to mock surgery.



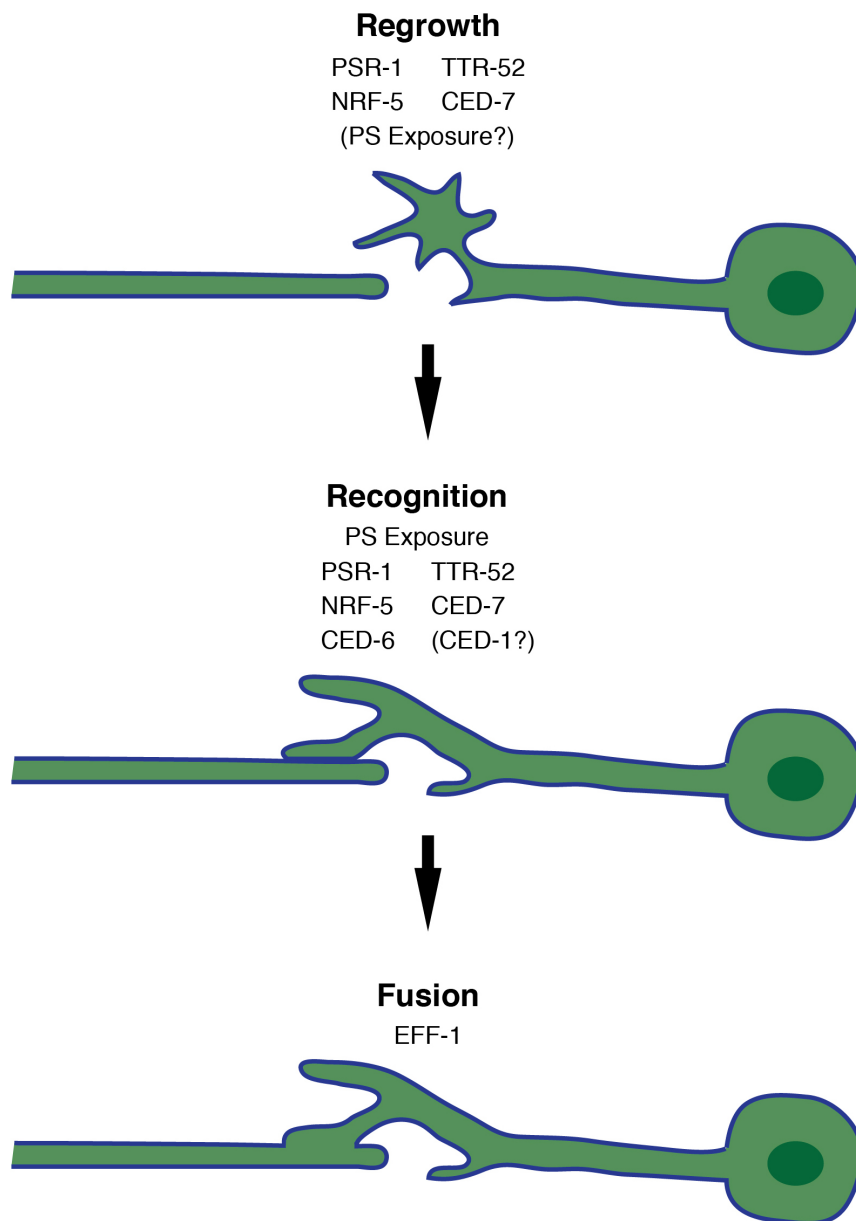
Extended Data Figure 8 | Mutation of *unc-70* causes axonal defects and PLM-PLN fusion events that are dependent on *eff-1* and *dlk-1*. **a**, In the absence of UNC-70 the PLM mechanosensory neurons develop abnormal morphology with breaks, loop structures (arrowhead) and branching. Scale bar, 25 μm ; image representative of 21 animals. **b**, Quantification of the penetrance of abnormal loop structures as well as axonal breakages; *n* value within the bar. **c**, To determine whether the fluorescence observed in PLN was a result of fusion with PLM, or due to a promoter-switching event, time-lapse imaging was undertaken *in vivo*, in *unc-70(s1502)* mutant animals. Single plane images captured on a spinning disk confocal are shown. Asterisk denotes PLM cell body (i); filled arrowhead indicates the site of initial GFP transfer first visible at 0 h (i); open arrowhead indicates site of eventual PLN cell body appearance, which occurs after GFP is visible in the axonal process (ii–ix). Scale bar, 50 μm . Image representative of one fusion event captured from 23 animals imaged. **d**, To further confirm the fusion events in *unc-70(s1502)* mutants, a tethered fluorophore assay was used to analyse cytoplasmic movement between PLM and PLN. Cytoplasmic GFP was expressed together with mRFP targeted to the mitochondria (TOMM-20::mRFP), both expressed

under the control of the *Pmec-4* promoter in the same extra-chromosomal array. In wild-type (WT) animals, both GFP and TOMM-20::mRFP are only visible in the soma and axon of the PLM neuron (no extra cell was visible). In those *unc-70(s1502)* mutant animals that presented PLM and PLN spontaneous neuronal fusion visible as GFP in both cells, PLM appeared as per wild-type with TOMM-20::mRFP in the soma and axon (left bar), whereas in PLN (designated 'Extra' in the bar) mitochondria were never observed. Animals with mutations in *sem-4(n1971)* and *ced-3(n717)* contain extra cells due to cell fate defects³⁵; as a result, both GFP and TOMM-20::mRFP are visible in the axon and soma of both PLM and the extra cells. Error bars represent standard error of proportion; *n* values for cells analysed in each bar. **e**, Spontaneous fusion between PLM and PLN is completely abolished in *dlk-1* mutants, and shows a large reduction in an *eff-1* mutant background. Overexpression of EFF-1 in the PLM neuron causes a large enhancement in spontaneous fusion events, but in animals lacking DLK-1, overexpression of EFF-1 in PLM is not sufficient to induce neuronal fusion between PLM and PLN. For **d** and **e**, error bars represent standard error of proportion. *n* values are shown adjacent to each bar. *P* values from *t*-test: ****P* < 0.001.



Extended Data Figure 9 | TTR-52 and sAnxV bind to the PLM axon in *unc-70* mutants. Localization of TTR-52::mCherry to the PLM axon at the L4 stage (a) or one-day-old adult (1DOA) animal (b) of *unc-70* mutant animals (genotype: *unc-70(s1502); oxIs95; smls119; zdlIs5*). Scale bars, 20 μ m. Images representative of 16 (a) and 17 (b) animals. c, Quantification of the relative fluorescence levels of TTR-52::mCherry on the PLM axons in L4 and 1DOA wild-type (WT) animals compared to *unc-70(s1502)* mutants. The intensity of TTR-52::mCherry on the axon shows a small but significant increase in *unc-70(s1502)* mutants compared to WT animals at the L4 stage and a more pronounced enhancement in 1DOA animals; $n = 15$ for WT L4, $n = 16$ for *unc-70* L4, and $n = 17$ for both genotypes at 1DOA stage. d, Localization of the

PS sensor, sAnxV::mRFP, to the PLM axon in an *unc-70(s1502)* mutant animal at the L4 stage. e, Localization of sAnxV::mRFP to the PLM axon in a 1DOA *unc-70(s1502)* mutant animal. Scale bars, 20 μ m. Images representative of 14 (d) and 15 (e) animals. f, Quantification of the relative fluorescence levels of sAnxV::mRFP on the PLM axons in L4 and 1DOA WT animals compared to *unc-70(s1502)* mutants. In 1DOA animals there is a large increase in the relative fluorescence in *unc-70(s1502)* mutants compared to WT animals; $n = 15$ for WT L4, $n = 14$ for *unc-70* L4, $n = 14$ for WT 1DOA, and 15 for *unc-70* 1DOA. For c and f, error bars indicate standard error. P values from Kolmogorov-Smirnov test: $*P < 0.05$, $***P < 0.001$. ns, not significant. au, arbitrary units.



Extended Data Figure 10 | Proposed model for how the apoptotic recognition molecules and the fusogen EFF-1 mediate axonal fusion during regeneration of PLM.

Allosteric activation of the RNF146 ubiquitin ligase by a poly(ADP-ribosyl)ation signal

Paul A. DaRosa^{1,2*}, Zhizhi Wang^{2*}, Xiaomo Jiang³, Jonathan N. Pruneda^{1†}, Feng Cong³, Rachel E. Klevit¹ & Wenqing Xu²

Protein poly(ADP-ribosyl)ation (PARylation) has a role in diverse cellular processes such as DNA repair, transcription, Wnt signalling, and cell death^{1–6}. Recent studies have shown that PARylation can serve as a signal for the polyubiquitination and degradation of several crucial regulatory proteins, including Axin and 3BP2 (refs 7–9). The RING-type E3 ubiquitin ligase RNF146 (also known as Iduna) is responsible for PARylation-dependent ubiquitination (PARdU)^{10–12}. Here we provide a structural basis for RNF146-catalysed PARdU and how PARdU specificity is achieved. First, we show that *iso*-ADP-ribose (*iso*-ADPr), the smallest internal poly(ADP-ribose) (PAR) structural unit, binds between the WWE and RING domains of RNF146 and functions as an allosteric signal that switches the RING domain from a catalytically inactive state to an active one. In the absence of PAR, the RING domain is unable to bind and activate a ubiquitin-conjugating enzyme (E2) efficiently. Binding of PAR or *iso*-ADPr induces a major conformational change that creates a functional RING structure. Thus, RNF146 represents a new mechanistic class of RING E3 ligases, the activities of which are regulated by non-covalent ligand binding, and that may provide a template for designing inducible protein-degradation systems. Second, we find that RNF146 directly interacts with the PAR polymerase tankyrase (TNKS). Disruption of the RNF146–TNKS interaction inhibits turnover of the substrate Axin in cells. Thus, both substrate PARylation and PARdU are catalysed by enzymes within the same protein complex, and PARdU substrate specificity may be primarily determined by the substrate–TNKS interaction. We propose that the maintenance of unliganded RNF146 in an inactive state may serve to maintain the stability of the RNF146–TNKS complex, which in turn regulates the homeostasis of PARdU activity in the cell.

Two domains can be identified within the 358-residue sequence of RNF146: an amino-terminal RING domain, followed by a WWE domain that binds *iso*-ADPr¹³ (Fig. 1a and Extended Data Fig. 1). RING E3 ligases activate a ubiquitin-conjugating enzyme (E2) to transfer ubiquitin directly from the E2 active site to a lysine residue of a substrate. The intrinsic ability of RING E3s to stimulate ubiquitin (Ub) transfer can be assayed by following Ub transfer from E2~Ub (activated complex with a thioester linkage) to free amino acid lysine¹⁴. Unexpectedly, neither the purified RNF146(RING) domain nor full-length RNF146 significantly enhance the rate of Ub transfer from UbCH5~Ub to lysine (Fig. 1b and Extended Data Fig. 2a). However, addition of the WWE ligands *iso*-ADPr or PAR to the full-length RNF146 or a fragment containing the RING and WWE domains (termed RNF146(RING-WWE)), but not a RING-only construct, results in robust activation (Fig. 1 and Extended Data Fig. 2), consistent with reports that RNF146 is a PARylation-dependent E3 (refs 10–12). Isothermal titration calorimetry (ITC) analysis shows that *iso*-ADPr binds to RNF146(RING-WWE) ten times tighter than to the WWE domain alone (dissociation constant (K_d) values of 39 nM and 372 nM, respectively; Extended Data Fig. 3a, b), suggesting that the presence of the RING domain contributes to ligand binding. Furthermore, RNF146(RING-WWE) is sensitive to subtilisin digestion in the absence

of *iso*-ADPr, but is more resistant to proteolysis in the presence of *iso*-ADPr (Extended Data Fig. 3c), and NMR experiments indicate structural changes within the RING domain of the RING-WWE fragment after *iso*-ADPr binding (Extended Data Fig. 4). Altogether, the data indicate that both the RING and WWE domains are involved in *iso*-ADPr binding and that ligand binding affects the conformation and/or stability of the RING domain, leading to increased E3 ligase activity.

The structural basis of RNF146 activation by *iso*-ADPr or PAR is evident in a 1.9 Å crystal structure of a RNF146(RING-WWE)–UbCH5a–*iso*-ADPr complex (Fig. 2 and Extended Data Table 1). The RING domain structure is largely similar to other structurally characterized RINGs (see below), and the RNF146(WWE) domain structure is almost identical to an existing crystal structure¹³ (Extended Data Fig. 5a). The most notable feature is the location of *iso*-ADPr, which contacts both the RING and WWE domains (Fig. 2b and Extended Data Fig. 5b, c). Contacts between the WWE domain and *iso*-ADPr are similar to those previously described¹³.

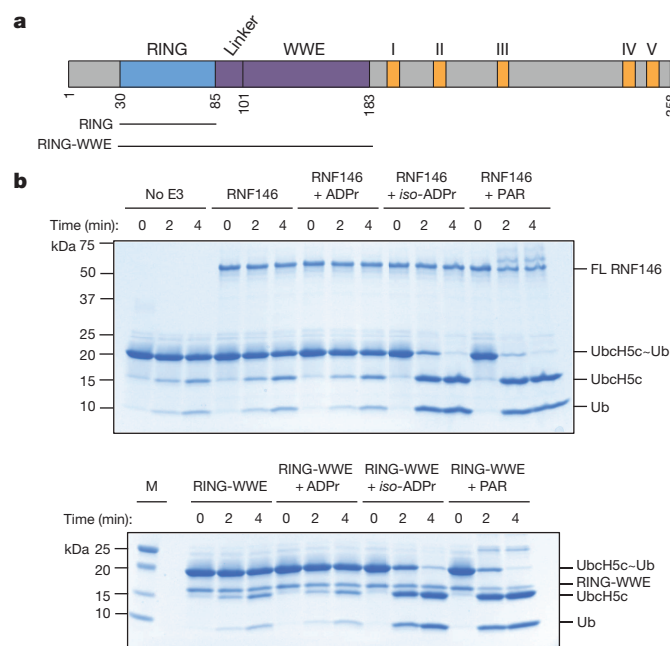


Figure 1 | Binding of *iso*-ADPr or PAR activates RNF146 E3 ubiquitin ligase activity. **a**, Domain structure of RNF146, with constructs used in this study shown: RNF146(RING), residues 30–89; RNF146(RING-WWE), residues 30–183. Five potential TNKS-binding motifs are shown in orange and numbered I to V. **b**, Coomassie-stained E2~Ub/lysine reactivity assays of full-length (FL) RNF146 (top) and RNF146(RING-WWE) (bottom) with the E2 UbCH5c. Active RING domains enhance the reactivity of E2~Ub with lysine and therefore speed the disappearance of the E2~Ub species with the corresponding appearance of free E2 and Ub. The WWE domain does not bind ADP-ribose (ADPr)¹³. M, molecular mass marker.

¹Department of Biochemistry, University of Washington, Seattle, Washington 98195, USA. ²Department of Biological Structure, University of Washington, Seattle, Washington 98195, USA. ³Novartis Institutes for Biomedical Research, Cambridge, Massachusetts 02139, USA. [†]Present address: Medical Research Council Laboratory of Molecular Biology, Cambridge CB2 0QH, UK.

*These authors contributed equally to this work.

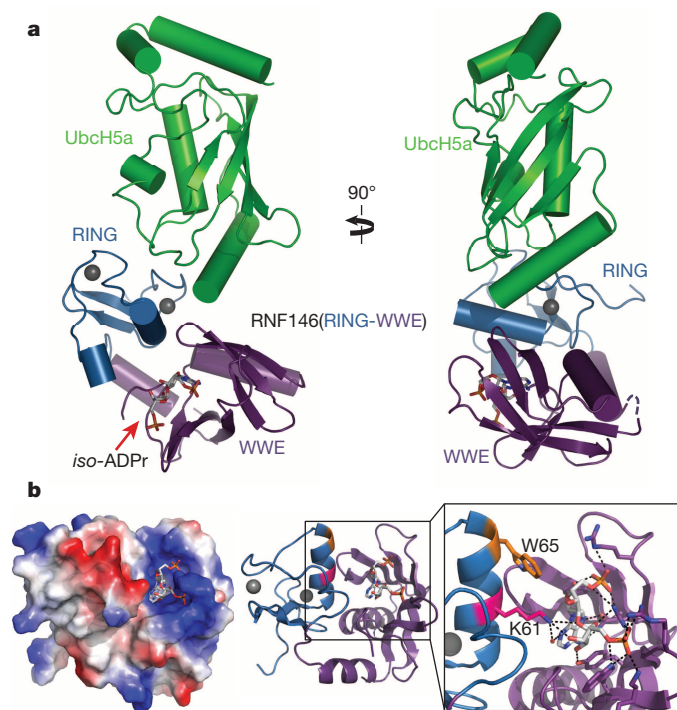


Figure 2 | Crystal structure of the RNF146(RING-WWE)-UbcH5a-iso-ADPr complex. **a**, Cartoon representation of the RNF146-UbcH5a complex with RING domain coloured blue, WWE domain coloured purple, and UbcH5a coloured green. Zn^{2+} ions are shown as grey spheres, and the *iso*-ADPr ligand is represented as sticks. **b**, The RNF146-*iso*-ADPr interface. Left, surface electrostatic view of RNF146(RING-WWE), showing the *iso*-ADPr/PAR-binding pocket; centre, same view, cartoon representation; right, close-up view of *iso*-ADPr pocket. Polar contacts between protein and the ligand, *iso*-ADPr, are indicated by dashed lines; RING residues Lys 61 (magenta) and Trp 65 (orange) are highlighted.

Lys 61 from the RING domain is within hydrogen-bond distance to hydroxyl groups on both ribose moieties of *iso*-ADPr and to a water molecule that can mediate a hydrogen bond with the adenine ring. In addition, RING residue Trp 65 forms a hydrogen bond with a main-chain carbonyl in the WWE domain and has van der Waals contacts with the ligand. Although *iso*-ADPr is buried in a valley between the RING and WWE domains, the phosphate groups on either end of *iso*-ADPr are exposed. Thus, the observed ligand orientation is consistent with the notion that RNF146 binds an internal unit of a PAR polymer.

In the crystal, the E2 enzyme UbcH5a binds to the RING domain at the canonical E2-E3 interface, away from the *iso*-ADPr-binding site (Fig. 2a). Similar to other RING E3-E2 complex structures^{15–24}, two Zn^{2+} -binding loops and the central helix of the RING bind E2 loops 4 and 7. A slight difference in the orientation of the RING and E2 relative to other E2-E3 complex structures is probably due to crystal packing (Extended Data Fig. 6), as the E2-E3 interactions observed in solution by NMR are similar to other well-characterized systems (see below).

Insight into the conformational changes that accompany *iso*-ADPr binding is provided by comparison to an NMR structure of the unliganded RNF146(RING) domain (Protein Data Bank (PDB) accession 2D8T; RIKEN Structural Genomics/Proteomics Initiative). In the unliganded RING domain the central helix is one turn shorter, with residues 62–66 instead forming a loop that protrudes into the E2-E3 binding interface (Fig. 3a and Extended Data Fig. 7a). Trp 65 makes hydrophobic interactions with Ile 36, Leu 66, and Ala 71 and is in a position to block E2 binding. Residues 62–66 adopt the helical structure associated with active RING domains in the *iso*-ADPr-bound structure. Thus, the RNF146(RING) domain can adopt two different conformations, and binding of *iso*-ADPr stabilizes an active structure with a functional E2-binding surface.

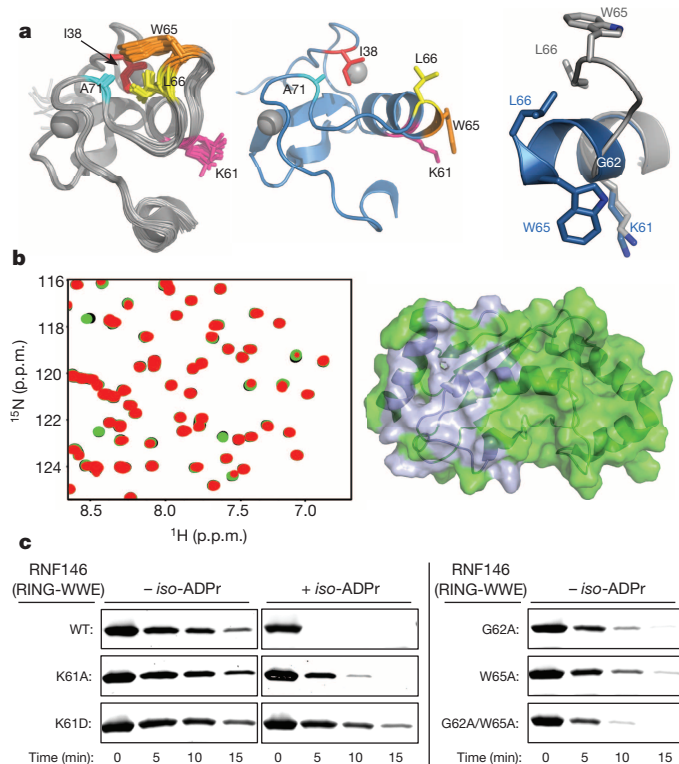


Figure 3 | Mechanism of RNF146 PAR-mediated RING activation. **a**, Left, apo-RNF146(RING) domain solution structure ensemble (residues 30–85; PDB accession 2D8T). Ile 38, Leu 66, Ala 71, and Trp 65 form a hydrophobic cluster in all members of the ensemble. Middle, the RING domain of the RNF146(RING-WWE)-UbcH5a-*iso*-ADPr complex (blue) adopts a canonical RING structure shown in the same orientation as the structure in the left panel. Right, helix 1 of the RNF146(RING) domain in our complex aligned with a representative NMR structure. After *iso*-ADPr binding, helix 1 is extended following Gly 62, and Trp 65 undergoes a marked relocation. **b**, Left, ^1H - ^{15}N heteronuclear single quantum coherence (HSQC) of ^{15}N -UbcH5c(Ser22Arg/Cys85Ser) alone (black), with 0.25 molar equivalence (m.e.) RNF146(RING-WWE) (green), and with 0.25 m.e. RNF146(RING-WWE) plus 0.5 m.e. of *iso*-ADPr (red). Right, chemical shift perturbations of residues in left panel mapped to the surface of UbcH5c (PDB accession 2FUH) show the binding surface for RNF146(RING-WWE) (light blue, on green E2). **c**, E2~Ub/lysine reactivity assays of RNF146(RING-WWE) mutants with the E2 UbcH5c; full gels are shown in Extended Data Fig. 8a.

The proposed model is supported by NMR binding experiments. Addition of unliganded RNF146(RING-WWE) resulted in very minor perturbations to the NMR spectrum of ^{15}N -UbcH5c while extensive perturbations occurred when *iso*-ADPr was also present (Fig. 3b and Extended Data Fig. 7b). Furthermore, the observed perturbations are highly similar to those observed with other RING-type E3s binding to UbcH5 and to UbcH5~Ub^{21,25} (Fig. 3b and Extended Data Fig. 7c, d).

Mutational analysis was performed to understand the function of key residues in the *iso*-ADPr-induced conformational switch. Mutation of RING Lys 61, which makes several contacts with *iso*-ADPr, to Ala or Asp increased the K_d for *iso*-ADPr to values comparable to that of the WWE domain alone (K_d values of 214 nM (RING-WWE(Lys61Ala)) or 457 nM (RING-WWE(Lys61Asp)) versus 372 nM (WWE); Extended Data Fig. 3a, b). Although the Lys61Asp mutant can still bind ligand, it is not activated by *iso*-ADPr (Fig. 3c and Extended Data Fig. 8a). Thus, Lys 61 serves to couple ligand binding to the activation of the RING domain. RING Gly 62 may serve to maintain the inactive RING conformation by disrupting the central helix (Extended Data Fig. 8b). Mutation of Gly 62 to Ala in the context of both the RNF146(RING) and RNF146(RING-WWE) constructs was performed. In the absence of ligand, the mutants promote E2~Ub lysine reactivity (Fig. 3c and Extended

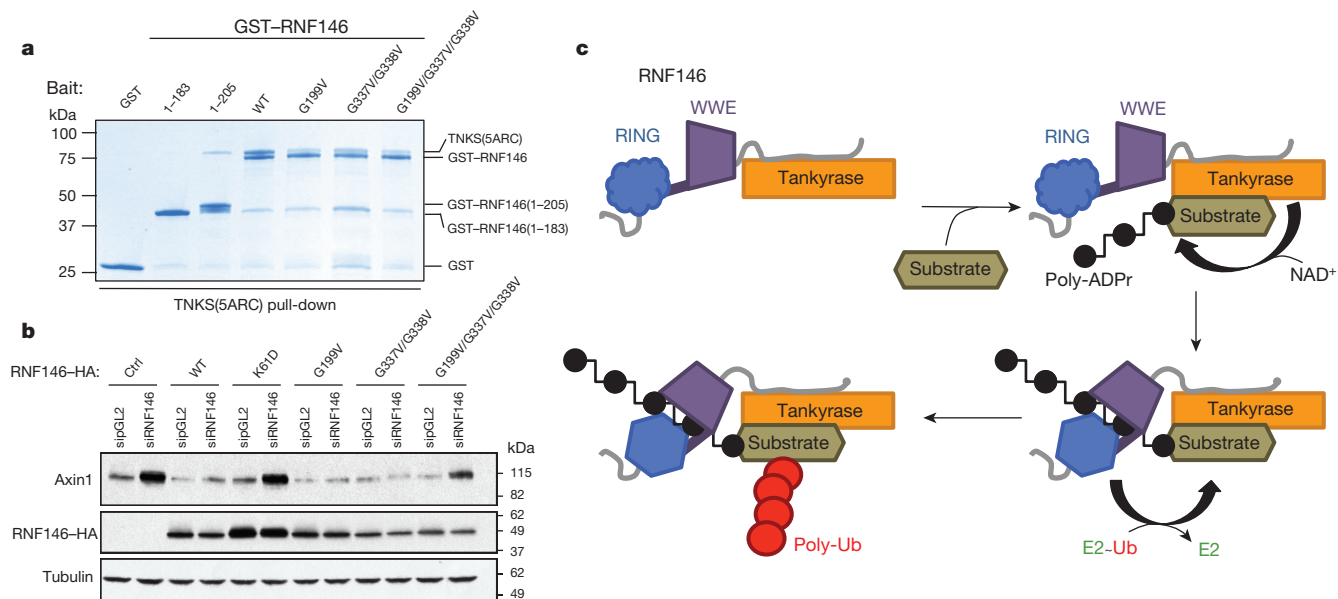


Figure 4 | RNF146 and TNKS form a tight complex crucial to PARdU *in vivo*. **a**, GST pull-downs of GST-tagged RNF146 variants with untagged TNKS(5ARC) (residues 173–961), demonstrate a direct interaction between RNF146 and the five ARCs of TNKS. The interaction probably involves several TNKS-binding sites in RNF146 as various RNF146 mutations reduce but do not abolish TNKS binding (inputs are shown in Extended Data Fig. 9d). **b**, Immunoblots of Axin turnover rescue assay show that both the TNKS–RNF146 interaction and the RNF146 allosteric switch are important for

Data Fig. 8c). RNF146(RING)(Gly62Ala) also shows increased E2 binding in NMR experiments (Extended Data Fig. 8d). Thus, Gly 62 may have a key role in the conformational transition of the central helix. Likewise, a Trp65Ala mutation in RNF146(RING–WWE) to disrupt Trp 65 interactions in the inactive state also increased basal E3 activity (Fig. 3c). The double mutant Gly62Ala/Trp65Ala of RNF146(RING–WWE) exhibits still greater activity than either of the single mutants (Fig. 3c). The mutational results are consistent with our model in which extension of the RING central helix and repositioning of Trp 65 from the E2–E3-binding site to the RING–*iso*-ADPr interface constitute the allosteric switch triggered by ligand binding.

Almost all known proteins regulated by PARdU, including Axin and 3BP2, are PARYlated by TNKS^{8,10–12}. We sought to understand the specificity of RNF146 by testing for a direct interaction between RNF146 and TNKS. Both glutathione *S*-transferase (GST) pull-down assays and co-migration on size-exclusion chromatography (SEC) showed that full-length RNF146 forms a stable complex with the five ankyrin repeat clusters (ARCs) of TNKS (TNKS(5ARC), residues 173–961) (Fig. 4a and Extended Data Fig. 9a). Co-immunoprecipitation and pull-down assays using full-length TNKS further support the direct RNF146–TNKS interaction in cells (Extended Data Fig. 9b, c).

The ARCs of TNKS recognize a consensus motif of RXXGDG^{6,9}, although our recent work suggested that deviations may be tolerated in some circumstances²⁶. As RNF146 contains no consensus motifs, we used molecular modelling to identify potential TNKS-binding sequences. Five potential motifs reside in the carboxy-terminal region of RNF146 (motif I to V; Fig. 1a and Extended Data Fig. 1); motif I (residues 193–199) is the most conserved. A C-terminal truncation (RNF146(1–183)) could not bind to the ARCs of TNKS, but a construct that includes motif I (RNF146(1–205)) bound TNKS(5ARC) detectably (Fig. 4a). In the context of full-length RNF146, mutations in either motif I (Gly199Val) or motif IV (Gly337Val/Gly338Val) reduced but did not abrogate the observed binding to TNKS(5ARC), with the motif I mutation having a stronger effect (Fig. 4a).

Knockdown of RNF146 in cells by short interfering RNA (siRNA) leads to increased steady-state levels of Axin, and this effect is reversed by

PARdU in cells. sipGL2, non-targeted siRNA control; siRNF146, siRNA targeting RNF146. **c**, Proposed TNKS–RNF146 PARdU model. RNF146 is inactive when bound to non-PARYlated TNKS in the cell. After substrate binding to TNKS and subsequent PARYlation, RNF146 binds an internal unit of PAR. This causes a conformational change in the RING domain, which activates its ligase activity, enabling the polyubiquitination of substrate. HA, haemagglutinin tag.

RNF146 expression^{7,11} (Fig. 4b). By contrast, expression of the allosteric-switch mutant RNF146(Lys61Asp) resulted in similar Axin levels to the no-RNF146 control. Disruption of the RNF146–TNKS complex via expression of the triple mutant RNF146(Gly199Val/Gly337Val/Gly338Val) also led to increased Axin levels (Fig. 4b). Thus, coupling of PAR binding to the stimulation of E3 activity, and the ability to form an RNF146–TNKS complex are both crucial to cellular PARdU regulation and turnover of Axin.

In summary, we propose that RNF146 and TNKS exist as a complex in which the RING domain is predominantly in an inactive state (Fig. 4c). TNKS is responsible for substrate selection and PARYlation. Subsequent ubiquitination of the PARYlated substrate requires that an internal PAR moiety binds to RNF146 to trigger the allosteric switch to the RING E3-active state. PARYlated substrate is probably held in the RNF146–TNKS complex via its interaction with TNKS, as the RNF146–TNKS interaction is required for ubiquitination of substrate in cells. Thus, our studies have revealed specific insights into the regulatory and substrate-recruitment mechanism of PARdU and have defined the molecular mechanism by which the RNF146(RING) domain is allosterically switched by non-covalent small molecule binding. These insights may aid in the design of RNF146 inhibitors that may be useful for cancer treatment, as RNF146 overexpression is associated with lung cancer²⁷. Furthermore, data presented here may provide insight into studies of PAR-dependent cell death (PARthanatos), in which RNF146 was identified to be important in protecting cells from PARthanatos^{12,28}.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 10 February; accepted 1 September 2014.

Published online 19 October 2014.

- Gibson, B. A. & Kraus, W. L. New insights into the molecular and cellular functions of poly(ADP-ribose) and PARPs. *Nature Rev. Mol. Cell Biol.* **13**, 411–424 (2012).
- Hottiger, M. O., Hassa, P. O., Luscher, B., Schuler, H. & Koch-Nolte, F. Toward a unified nomenclature for mammalian ADP-ribosyltransferases. *Trends Biochem. Sci.* **35**, 208–219 (2010).

3. Luo, X. & Kraus, W. L. On PAR with PARP: cellular stress signaling through poly(ADP-ribose) and PARP-1. *Genes Dev.* **26**, 417–432 (2012).
4. Wang, Y., Dawson, V. L. & Dawson, T. M. Poly(ADP-ribose) signals to mitochondrial AIF: a key event in parthanatos. *Exp. Neurol.* **218**, 193–202 (2009).
5. Virág, L. 50 Years of poly(ADP-ribosylation). *Mol. Aspects Med.* **34**, 1043–1045 (2013).
6. Hsiao, S. J. & Smith, S. Tankyrase function at telomeres, spindle poles, and beyond. *Biochimie* **90**, 83–92 (2008).
7. Huang, S. M. *et al.* Tankyrase inhibition stabilizes axin and antagonizes Wnt signalling. *Nature* **461**, 614–620 (2009).
8. Levaot, N. *et al.* Loss of Tankyrase-mediated destruction of 3BP2 is the underlying pathogenic mechanism of cherubism. *Cell* **147**, 1324–1339 (2011).
9. Guettler, S. *et al.* Structural basis and sequence rules for substrate recognition by Tankyrase explain the basis for cherubism disease. *Cell* **147**, 1340–1354 (2011).
10. Callow, M. G. *et al.* Ubiquitin ligase Rnf146 regulates Tankyrase and axin to promote Wnt signaling. *PLoS ONE* **6**, e22595 (2011).
11. Zhang, Y. *et al.* RNF146 is a poly(ADP-ribose)-directed E3 ligase that regulates axin degradation and Wnt signalling. *Nature Cell Biol.* **13**, 623–629 (2011).
12. Kang, H. C. *et al.* Iduna is a poly(ADP-ribose) (PAR)-dependent E3 ubiquitin ligase that regulates DNA damage. *Proc. Natl Acad. Sci. USA* **108**, 14103–14108 (2011).
13. Wang, Z. *et al.* Recognition of the iso-ADP-ribose moiety in poly(ADP-ribose) by WWE domains suggests a general mechanism for poly(ADP-ribosylation)-dependent ubiquitination. *Genes Dev.* **26**, 235–240 (2012).
14. Wenzel, D. M., Lissounov, A., Brzovic, P. S. & Klevit, R. E. UBCH7 reactivity profile reveals parkin and HHAR1 to be RING/HECT hybrids. *Nature* **474**, 105–108 (2011).
15. Zheng, N., Wang, P., Jeffrey, P. D. & Pavletich, N. P. Structure of a c-Cbl-UbcH7 complex: RING domain function in ubiquitin-protein ligases. *Cell* **102**, 533–539 (2000).
16. Das, R. *et al.* Allosteric regulation of E2:E3 interactions promote a processive ubiquitination machine. *EMBO J.* **32**, 2504–2516 (2013).
17. Dou, H. *et al.* Structural basis for autoinhibition and phosphorylation-dependent activation of c-Cbl. *Nature Struct. Mol. Biol.* **19**, 184–192 (2012).
18. Bentley, M. L. *et al.* Recognition of UbcH5c and the nucleosome by the Bmi1/Ring1b ubiquitin ligase complex. *EMBO J.* **30**, 3285–3297 (2011).
19. Campbell, S. J. *et al.* Molecular insights into the function of RING finger (RNF)-containing proteins hRNF8 and hRNF168 in Ubc13/Mms2-dependent ubiquitylation. *J. Biol. Chem.* **287**, 23900–23910 (2012).
20. Yin, Q. *et al.* E2 interaction and dimerization in the crystal structure of TRAF6. *Nature Struct. Mol. Biol.* **16**, 658–666 (2009).
21. Pruneda, J. N. *et al.* Structure of an E3:E2~Ub complex reveals an allosteric mechanism shared among RING/U-box ligases. *Mol. Cell* **47**, 933–942 (2012).
22. Plechanovová, A., Jaffray, E. G., Tatham, M. H., Naismith, J. H. & Hay, R. T. Structure of a RING E3 ligase and ubiquitin-loaded E2 primed for catalysis. *Nature* **489**, 115–120 (2012).
23. Hodson, C., Purkiss, A., Miles, J. A. & Walden, H. Structure of the human FANCL RING-Ube2T complex reveals determinants of cognate E3–E2 selection. *Structure* **22**, 337–344 (2014).
24. Dou, H., Buetow, L., Sibbet, G. J., Cameron, K. & Huang, D. T. BIRC7–E2 ubiquitin conjugate structure reveals the mechanism of ubiquitin transfer by a RING dimer. *Nature Struct. Mol. Biol.* **19**, 876–883 (2012).
25. Brzovic, P. S. *et al.* Binding and recognition in the assembly of an active BRCA1/BARD1 ubiquitin-ligase complex. *Proc. Natl Acad. Sci. USA* **100**, 5646–5651 (2003).
26. Morrone, S., Cheng, Z., Moon, R. T., Cong, F. & Xu, W. Crystal structure of a Tankyrase-Axin complex and its implications for Axin turnover and Tankyrase substrate recruitment. *Proc. Natl Acad. Sci. USA* **109**, 1500–1505 (2012).
27. Gao, Y. *et al.* Overexpression of RNF146 in non-small cell lung cancer enhances proliferation and invasion of tumors through the Wnt/β-catenin signaling pathway. *PLoS ONE* **9**, e85377 (2014).
28. Andrabi, S. A. *et al.* Iduna protects the brain from glutamate excitotoxicity and stroke by interfering with poly(ADP-ribose) polymer-induced cell death. *Nature Med.* **17**, 692–699 (2011).

Acknowledgements We thank P. Brzovic and N. Zheng for discussions and editorial comments. We are grateful to the staff at Advanced Light Source (ALS) beamlines BL 8.2.1 and 8.2.2 for assistance with synchrotron data collection. This work was supported by National Institutes of Health (NIH) grant R01 GM099766 to W.X. and R.E.K. and NIH T32 GM07270 to P.A.D.

Author Contributions P.A.D. and Z.W. performed experiments. X.J. performed cell-based assays. F.C. and J.N.P. provided critical insight. P.A.D., Z.W., R.E.K. and W.X. wrote the paper. All authors provided editorial comments.

Author Information The atomic coordinates and structure factors of the RNF146-UbcH5a-iso-ADPr complex are deposited in the Protein Data Bank (PDB) with the accession code 4QPL. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to W.X. (wxu@uw.edu) and R.E.K. (klevit@uw.edu).

METHODS

Plasmids, protein expression and purification. Human RNF146 was subcloned into a pET-28a vector with an N-terminal His₆ and T7 tags and a tobacco etch virus (TEV) cleavage site, and a pGEX-6P-2 vector with an N-terminal GST and C-terminal His₆ tag. Mouse RNF146(RING-WWE) (residues 30–183), which has an identical protein sequence to human RNF146 in this region, was subcloned into a pGEX-4T-1 vector with an N-terminal GST tag and TEV cleavage site. GST-tagged human tankyrase 1 fragment containing all five ankyrin repeat clusters (TNKS(5ARC); residues 173–961) was expressed from a pGEX-4T-1 vector with an N-terminal GST tag and TEV cleavage site. Full-length mouse TNKS was cloned into a pET-15b plasmid. Mutants and truncations of RNF146 were generated using site-directed mutagenesis (Stratagene) and confirmed by sequencing. BRCA1–BARD1 (residues 1–112 and 1–115 respectively), UbcH5a, UbcH5b, UbcH5c, Ub, UbcH5c(Ser22Arg/Cys85Ser), and wheat E1 were purified as previously described^{25,29}. The oxyester-linked E2–O–Ub conjugate (UbcH5c(Ser22Arg/Cys85Ser)–O–Ub) was generated and purified as previously described³⁰.

All *Escherichia coli* (BL21) cultures were grown to an attenuation (*D*) at 600 nm of 0.6–1.2 in LB media or minimal MOPS media supplemented with ¹⁵N-ammonium chloride (Cambridge Isotope Labs) for NMR spectroscopy. Protein expression was induced in the presence of 200 μM isopropylthiogalactoside (IPTG) at 16 °C or 24 °C for 16–18 h. Bacterially expressed GST–RNF146(RING-WWE) or GST–RNF146(RING) (residues 30–89) was bound to glutathione sepharose 4B resin (GE Healthcare), washed, and eluted with 10 mM glutathione in a buffer containing 25 mM sodium phosphate, pH 7.6, and 200 mM NaCl. GST was then cleaved from proteins using a His₆–TEV protease for 1 h at 37 °C (or overnight at 4 °C) and the samples were dialysed overnight at 4 °C into 4 l of phosphate buffer (25 mM sodium phosphate, pH 7.6, 200 mM NaCl). Dialysed proteins were then run through Ni²⁺ NTA resin (GE Healthcare) to capture TEV and subsequently glutathione sepharose 4B resin to capture GST. After concentrating in the presence of 2 mM dithiothreitol (DTT), RNF146(RING-WWE) was purified using Superdex 75 (GE Healthcare) equilibrated with 25 mM phosphate, pH 7.0, 150 mM NaCl. Similarly, TNKS(5ARC) was purified using glutathione sepharose 4B resin, followed by on-column TEV cleavage overnight at 4 °C. The untagged TNKS(5ARC) was subsequently purified by anion exchange column using a Q column (GE Healthcare). Full-length mouse TNKS1 was partially purified by Ni²⁺ NTA resin (GE Healthcare). Full-length His₆–T7–RNF146 was purified using a Ni²⁺ NTA resin (GE Healthcare), followed by ion exchange on a HiTrap Q column (GE Healthcare) and SEC using a Superdex 200 10/300 column (GE Healthcare). Protein concentrations were determined by their ultraviolet absorbance at 280 nm, and confirmed with Coomassie-stained SDS–PAGE.

Lysine reactivity assay. For Coomassie or Oriole (BioRad)-stained gels, UbcH5~Ub conjugates were generated in a solution containing 100 μM E2, 1.5 μM wheat E1, 200 μM Ub (Lys-deficient (K0) mutant: Lys6Arg, Lys11Arg, Lys29Arg, Lys33Arg, Lys48Arg, Lys63Arg and Lys27Met), 2.5 mM MgCl₂ and 2 mM ATP (Sigma-Aldrich) in PBS at 37 °C for 30 min. The conjugate was purified by SEC before E2~Ub/lysine reactions. E2~Ub conjugates were added to E3 samples (wild-type or mutant full-length RNF146, RNF146(RING-WWE) or RNF146(RING)) for a final concentration of 25 μM E2 and 4 or 8 μM E3 (all reactions in Fig. 3c were performed at 8 μM E3), and 25 μM *iso*-ADPr where indicated. Time zero samples were taken just before the addition of buffered L-lysine HCl (Sigma-Aldrich) to a final concentration of 40 mM and incubated at 35 °C. Time points were collected between 1 and 15 min; the reactions were quenched in 2× non-reducing SDS loading buffer, and analysed by SDS–PAGE.

For western blot analysis, 5 μM UbcH5a enzyme was charged with 0.5 μM E1, 5 μM Ub (HA-tagged, K0 mutant: Lys6Arg, Lys11Arg, Lys29Arg, Lys33Arg, Lys48Arg, Lys63Arg and Lys27Met), 10 mM MgCl₂, and 5 mM ATP for 30–40 min at 37 °C. Reaction mixtures were added to tubes containing E3 for a final concentration of 5 μM E3 and 10 μM *iso*-ADPr when appropriate. Time zero samples were taken just before addition of buffered L-lysine HCl (Sigma-Aldrich) to a final concentration of 25 mM and incubated at 35 °C. Reactions were quenched in 2× non-reducing SDS loading buffer, analysed by western blots with mouse anti-HA (monoclonal antibody, Covance MMS-101P).

Auto-ubiquitination. *In vitro* auto-ubiquitination was performed in a reaction mixture containing 1 μM E1, 3 μM UbcH5c, 45 μM ubiquitin, 5 μM MgCl₂, and 3 μM His₆–T7–RNF146. PAR or *iso*-ADPr was added to a final concentration of about 20 μM (ADPr units) or 10 μM respectively. ATP was added to a final concentration of 5 μM after collecting time zero samples. The samples were then incubated at 37 °C and time points were collected between 0 and 12 min. Reactions were quenched with 10× reducing SDS loading buffer, boiled and analysed by western blot using mouse anti-T7 antibody (monoclonal antibody, Novagen 69522).

Limited proteolysis. Limited proteolysis was performed by incubating 10 μg RNF146(RING-WWE) (residues 30–183) or RNF146(Linker-WWE) (residues 83–183) with 50 ng subtilisin (Sigma-Aldrich) in the presence or absence of 150 μM

iso-ADPr for 1 h at room temperature. The reaction was quenched with 5× SDS loading buffer and the resulting digests were resolved by SDS–PAGE and stained with Coomassie.

SEC–MALS. SEC–MALS experiments were performed at room temperature with a Superdex 200 10/300 column (GE Healthcare) and a miniDAWN TREOS MALS detector (Wyatt Technology). About 200 μg of full-length RNF146, TNKS(5ARC), and TNKS(5ARC)–RNF146 complex was injected in each run. The column was run with buffer containing 20 mM Tris–HCl, pH 7.5, 150 mM NaCl, and 2 mM DTT. The light scattered by a protein is directly proportional to its weight-average molecular mass and concentration.

Crystallization of the RNF146(RING-WWE)–UbcH5a–*iso*-ADPr complex. RNF146(RING-WWE) was bound to an SP anion exchange column (GE Healthcare) and then eluted with buffer containing 20 mM Tris–HCl, pH 7.0, 5% (v/v) glycerol, 2 mM DTT, and 500 mM NaCl. The protein was eluted with a salt concentration of ~150 mM NaCl and was subsequently concentrated to 2.4 mg ml^{−1} for crystallization. RNF146(RING-WWE) was mixed with *iso*-ADPr ligand in a microcentrifuge tube for a final concentration of 133 μM protein to 267 μM ligand and incubated on ice for 30 min. Eighty microlitres of this solution was mixed with 20 μl of His₆–UbcH5a in a 25 mM sodium phosphate, pH 7.0, 150 mM NaCl buffer, for final concentrations of 120 μM E2 and 106 μM E3. This mixture was incubated on ice for 30 min. Crystals were obtained using the hanging drop method by adding 1 μl of the above protein–ligand mixture to 1 μl of a well solution containing 800 mM sodium citrate, 80 mM Tris–HCl, pH 7.0, 160 mM NaCl, 4 mM DTT, and 20 mM trimethylamine HCl. Crystals formed in 24–36 h at room temperature. Crystals were frozen with liquid nitrogen in a cryoprotectant composed of 800 mM sodium citrate, 96 mM Tris–HCl, pH 7.0, 280 mM NaCl, 5 mM DTT, and 20% (v/v) glycerol.

X-ray data collection and structure determination. Crystal screening and data collection were performed at the ALS, beamline 8.2.1. All diffraction data were processed by the HKL2000 package³¹ in the space group P2₁2₁2. The structure was determined, at 1.9 Å resolution, by single-wavelength anomalous dispersion (SAD) using one data set collected at a wavelength of 1.28295 Å, which was also used for refinement (Extended Data Table 1). The zinc sites and the initial phases were determined by PHENIX³². Four zinc sites were found in one asymmetric unit, and the experimental electron density map clearly showed the presence of two RNF146(RING-WWE)–UbcH5a complexes with two ligands in one asymmetric unit. The complex model was improved using iterative cycles of manual rebuilding with the program COOT³³ and refinement with Refmac5 of the CCP4 6.4.0 program suite³⁴. There are no Ramachandran outliers (96.0% most favoured, 4.0% allowed). The electrostatic potential surfaces shown were generated with the APBS tool in PyMol³⁵.

ITC. ITC was performed on a VP-ITC Microcal calorimeter (Microcal) at 25 °C for RNF146(RING-WWE) and the RNF146(RING-WWE) mutants. Protein and *iso*-ADPr were buffer exchanged into 20 mM HEPES, pH 7.5, 150 mM NaCl, and 1 mM DTT using SEC on a Superdex 75 10/300 column (GE Healthcare), eluting at a final concentration of ~15–20 μM. The ligand, *iso*-ADPr, was diluted in the same buffer to ~500 μM. Both ligand and protein were degassed before use. Ligand was injected in 5 μl quantities every 5 min for a total of 25–40 injections into a 1.4218-ml protein chamber. Data were analysed using Origin 7.0, fitting curves to a one-site model.

NMR spectroscopy. Two-dimensional ¹H–¹⁵N HSQC–TROSY experiments were performed on a Bruker 500 MHz AVANCE II NMR spectrometer. All data were obtained with 200 μM ¹⁵N-labelled protein. Data were processed with NMRPipe³⁶, and peak intensities (I) and chemical shift perturbations (CSPs) were measured in NMRViewJ³⁷ (One Moon Scientific). Peak intensity changes of ¹⁵N-labelled E2 (UbcH5c(Ser22Arg/Cys85Ser)) were measured relative to free E2 (*I*_{E3 bound}/*I*_{free}), and peaks affected by binding were identified by one standard deviation away from the mean *I*_{E3 bound}/*I*_{free} value. Chemical shift perturbations were determined by the equation $\Delta\delta_i = [(\Delta\delta_i^{15N}/5)^2 + (\Delta\delta_i^1H)^2]^{1/2}$; peaks with CSPs greater than one standard deviation from the mean were considered significantly perturbed. Both peak broadening and CSPs were mapped onto the E2 surface (PDB code 2FUH)²⁹ to show residues affected by E3 binding. For ¹H–¹⁵N HSQC–TROSY experiments of the oxyester-linked E2–O–Ub conjugate (UbcH5c(Ser22Arg/Cys85Ser)–O–Ub), Ub and UbcH5c peak intensities and CSPs were analysed separately as indicated above. For NMR experiments of the E2–O–Ub conjugate, peak broadening and CSPs were mapped onto the UbcH5b and Ub surfaces of the BIRC7–UbcH5b–O–Ub structure²⁴ (PDB accession 4AUQ) to show residues most affected when E3 binds.

GST pull-down assays. Approximately 3 μM of purified GST or GST–RNF146 (wild type, truncations or mutants) and about 6.7 μM of untagged TNKS(5ARC) were incubated with 25 μl of GSH sepharose 4B resin (GE Healthcare) for 1 h at room temperature in binding buffer (40 mM Tris–HCl, pH 8.0, 100 mM NaCl, 0.04% (v/v) Tween 20, and 2 mM DTT; total volume of 200 μl). The beads were then washed with 200 μl buffer three times. Proteins were resolved on SDS–PAGE and stained with Coomassie.

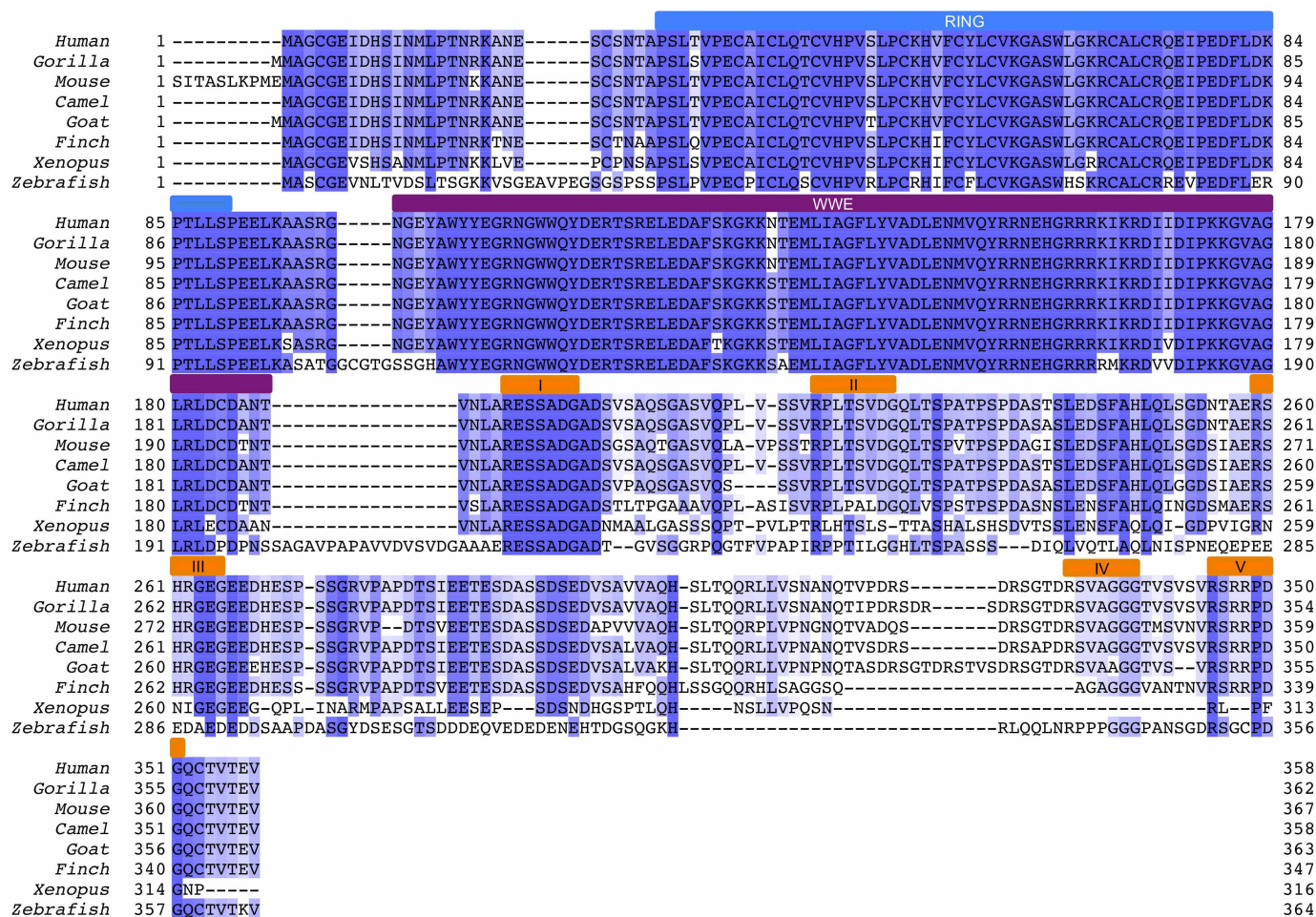
Full-length TNKS pull-down assays were performed using GST-RNF146 (Arg163Ala), a mutant that is deficient in binding poly(ADP-ribose)¹¹. Proteins were resolved via SDS-PAGE and the full-length TNKS, GST and GST-RNF146 (Arg163Ala) were visualized by western blot analysis using anti-TNKS (rabbit polyclonal antibody, Abcam ab86279) and anti-GST (mouse monoclonal antibody, GenScript A00865) antibodies.

Immunoprecipitation. HEK293T cells (ATTC, CRL-11268) were transfected with indicated plasmids with Fugene HD (Promega). Thirty-six hours after transfection, cells were lysed with immunoprecipitation buffer (20 mM Tris-HCl, pH 7.4, 150 mM NaCl, 1% (v/v) Triton X-100, supplemented with protease inhibitors and phosphatase inhibitors) and cleared by centrifugation. Flag-tagged proteins were immunoprecipitated from lysates with Flag-agarose beads (Sigma-Aldrich), and washed in immunoprecipitation buffer. Proteins bound to the beads were resolved by SDS-PAGE and analysed by immunoblotting. Cells were authenticated by single nucleotide polymorphism testing and confirmed as mycoplasma negative by a PCR-based assay.

RNF146 knockdown and rescue assay. siRNA-mediated knockdown and cDNA overexpression of RNF146 were described previously¹¹. siRNA-resistant HA-tagged RNF146 (wild type) and indicated mutants were subcloned into pcDNA4-TO. T-REx-293 cells (Life Technologies) were transfected with these plasmids and selected with Blasticidin to establish stable lines that express RNF146 in a doxycycline-inducible manner. siRNAs against RNF146 and luciferase (as negative control) were transfected with Lipofectamine RNAiMax (Life Technologies). Sequences of siRNAs used are: RNF146 sense, 5'-GCACGUUUUCUGCUAUCUAdTdT-3', antisense, 5'-UAGAUAGCAGAAAACGUGCdTdT-3' (Qiagen); pGL2 (luciferase), sense, 5'-CGUACGCGGAAUACUUCGAdTdT-3', antisense, 5'-UCGAAGUAUUCGCGUACGdTdT-3' (Dharmacon). Seventy-two hours after siRNA transfection and doxycycline induction, cells were lysed in RIPA buffer (50 mM Tris-HCl, pH 7.4, 150 mM NaCl, 1% (v/v) NP-40, 0.5% sodium deoxycholate, 0.1%

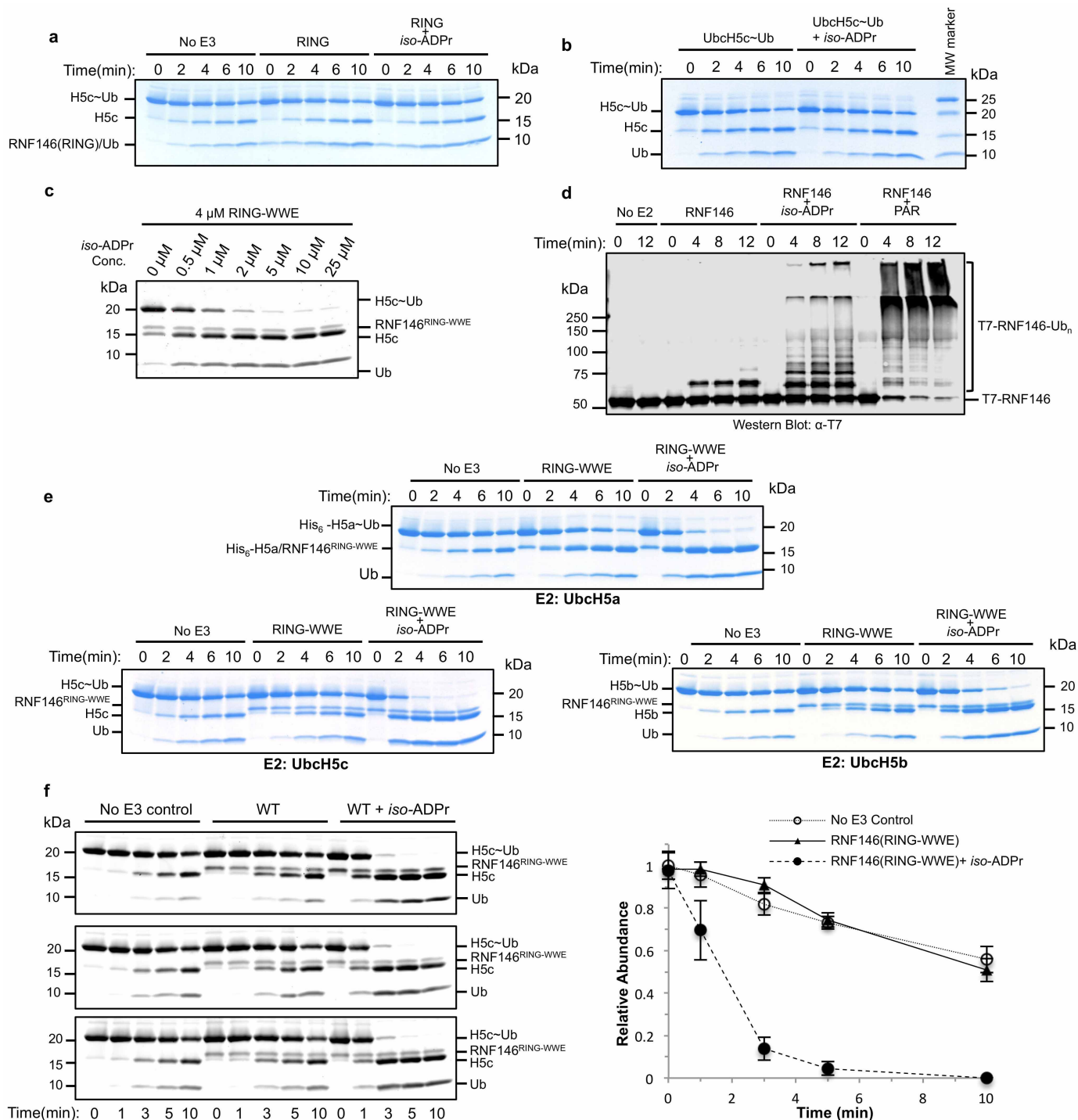
SDS, 1 mM EDTA) supplemented with protease inhibitor cocktail (Roche) and phosphatase inhibitor cocktail (Thermo Scientific). Cell lysates were resolved by SDS-PAGE and analysed by immunoblotting with rabbit anti-Axin1 (monoclonal antibody, Cell Signaling Technology 2075), rat anti-HA (monoclonal antibody, Roche 12158167001), and mouse anti-tubulin (monoclonal antibody, Sigma T5168). T-REx-293 cells were authenticated by single nucleotide polymorphism testing and confirmed as mycoplasma negative by a PCR-based assay.

29. Brzovic, P. S., Lissounov, A., Christensen, D. E., Hoyt, D. W. & Klevit, R. E. A. UbcH5/ubiquitin noncovalent complex is required for processive BRCA1-directed ubiquitination. *Mol. Cell* **21**, 873–880 (2006).
30. Pruneda, J. N., Stoll, K. E., Bolton, L. J., Brzovic, P. S. & Klevit, R. E. Ubiquitin in motion: structural studies of the ubiquitin-conjugating enzyme~ubiquitin conjugate. *Biochemistry* **50**, 1624–1633 (2011).
31. Otwinowski, Z. & Minor, W. *Processing of X-ray Diffraction Data Collected in Oscillation Mode* Vol. 276 (Academic, 1997).
32. Adams, P. D. et al. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D* **66**, 213–221 (2010).
33. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. D* **66**, 486–501 (2010).
34. The CCP4 suite. Programs for protein crystallography. *Acta Crystallogr. D* **50**, 760–763 (1994).
35. Delano, W. L. & Brunger, A. T. Helix packing in proteins: prediction and energetic analysis of dimeric, trimeric, and tetrameric GCN4 coiled coil structures. *Proteins* **20**, 105–123 (1994).
36. Delaglio, F. et al. NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J. Biomol. NMR* **6**, 277–293 (1995).
37. Johnson, B. A. & Blevins, R. A. NMR View: A computer program for the visualization and analysis of NMR data. *J. Biomol. NMR* **4**, 603–614 (1994).
38. Yeh, T. Y. et al. Tankyrase recruitment to the lateral membrane in polarized epithelial cells: regulation by cell-cell contact and protein poly(ADP-ribosylation). *Biochem. J.* **399**, 415–425 (2006).



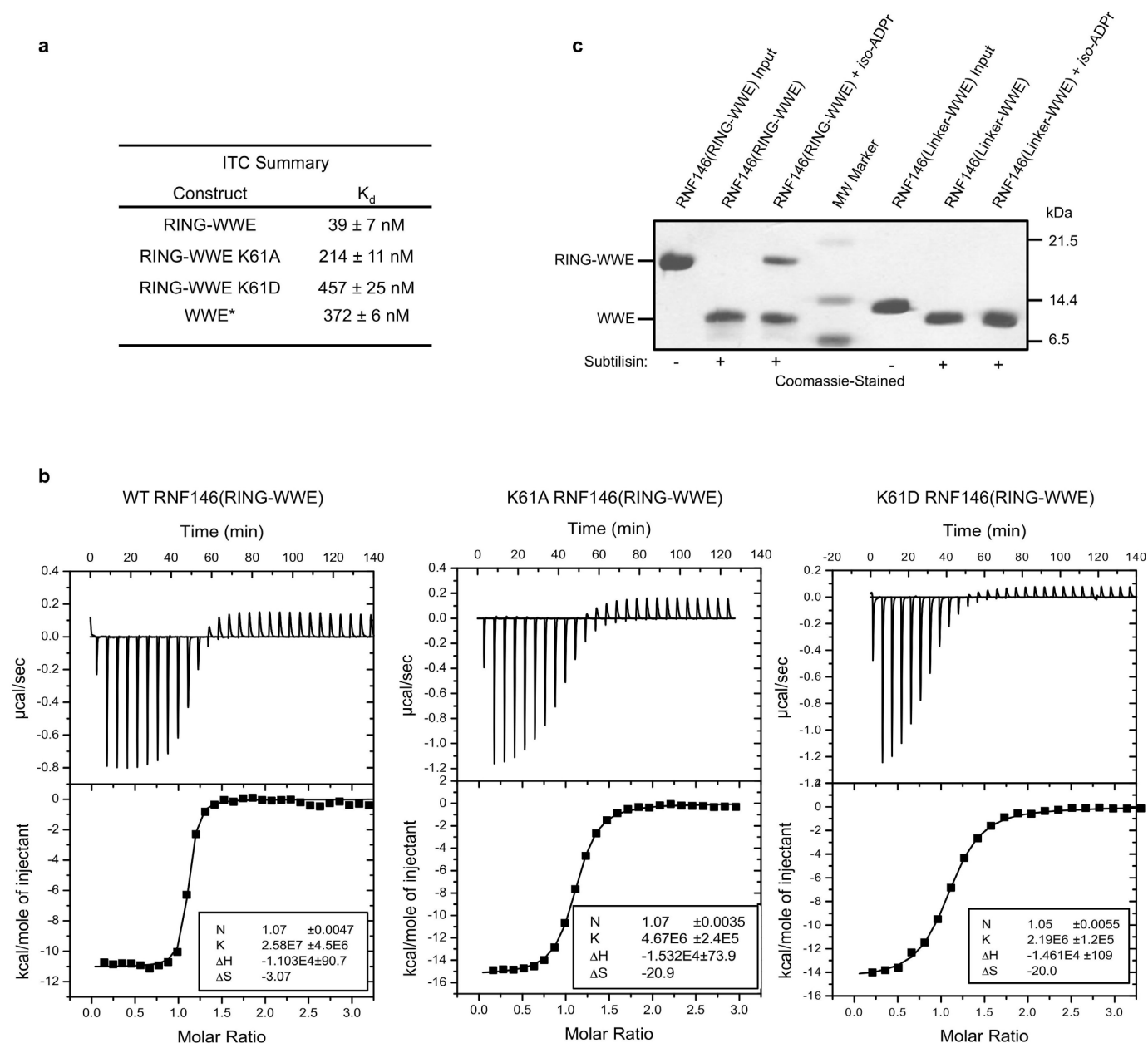
Extended Data Figure 1 | Multiple sequence alignment of RNF146 orthologues. The coloured bars above the sequence alignment indicate regions of interest in human RNF146: RING domain (blue), WWE domain (purple), and potential TNKS-binding motifs, numbered I to V (orange). Although there

are no apparent RXXGDG TNKS-binding motifs, the five potential binding motifs indicated here are based on the TNKS-substrate interface plasticity demonstrated by a recent crystal structure of the Axin-TNKS complex²⁶.



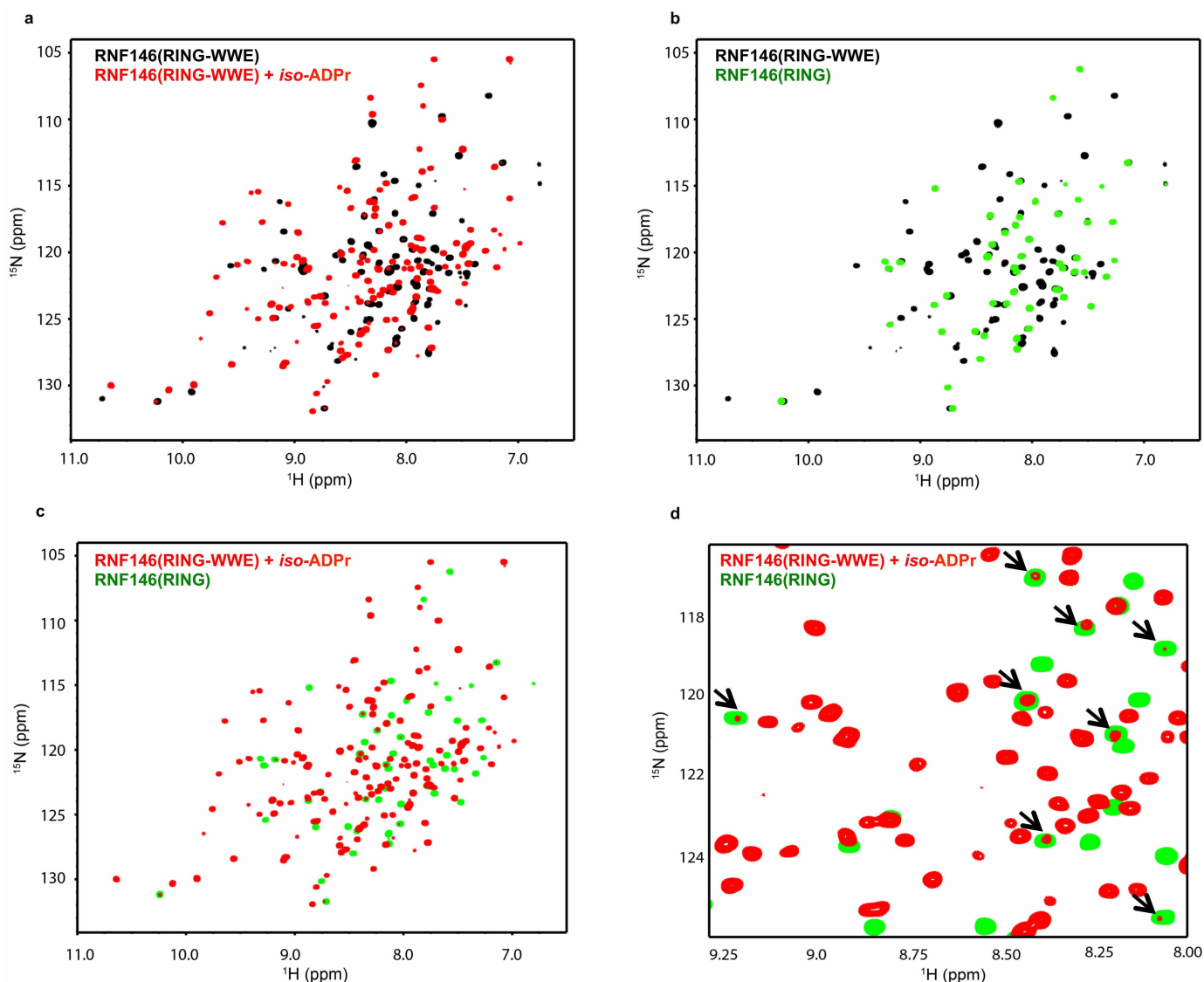
Extended Data Figure 2 | Both PAR and *iso*-ADPr can activate RNF146 E3 ligase activity. **a**, Coomassie-stained E2~Ub/lysine reactivity of the RNF146(RING) domain with and without *iso*-ADPr. The RING domain does not enhance E2~Ub conjugate reactivity in the absence or presence of ligand. **b**, Intrinsic lysine reactivity of the UbcH5c~Ub conjugate with and without *iso*-ADPr. *iso*-ADPr does not enhance the reactivity of the conjugate in the absence of RNF146. **c**, Oriole-stained E2~Ub/lysine reactivity with increasing *iso*-ADPr (3 min after lysine addition). The rate of E2~Ub/lysine reactivity is increased as a function of [*iso*-ADPr] up to 5 μ M ligand addition (1.2 equiv.), consistent with the affinity of RNF146 for *iso*-ADPr (see Extended Data Fig. 3). **d**, Auto-ubiquitination of full-length RNF146 in the absence or presence of *iso*-ADPr or PAR polymer. Image shows western blot for T7-tagged RNF146.

Because full-length RNF146 and the RING-WWE fragment have similar abilities to enhance E2~Ub reactivity (see Fig. 1), the additional auto-ubiquitination seen with PAR is probably due to increased local concentration of RNF146 near PAR polymers, allowing auto-ubiquitination *in trans*. **e**, E2/lysine reactivity of UbcH5a, UbcH5b, and UbcH5c ubiquitin conjugates with RNF146(RING-WWE) in the absence or presence of *iso*-ADPr (Coomassie-stained). All three isoforms function with ligand-activated RNF146. **f**, Technical triplicates of RNF146(RING-WWE) E2~Ub/lysine reactivity assays (Oriole-stained; left) and a plot of relative densitometry values of the E2~Ub conjugate (right). Error bars indicate the mean \pm s.d. from three separate experiments. All times are given in minutes. 'No E3' samples do not contain RNF146.



Extended Data Figure 3 | Both RNF146(RING) and RNF146(WWE) domains contribute to *iso*-ADPr binding. **a**, Summary of *iso*-ADPr binding (K_d values) for RNF146(RING-WWE) obtained from the ITC titrations in the current work, and for the WWE-only fragment (previously published; indicated by asterisk)¹³. These data indicate that the RING domain contributes to *iso*-ADPr binding. **b**, Raw ITC titrations of RNF146(RING-WWE)

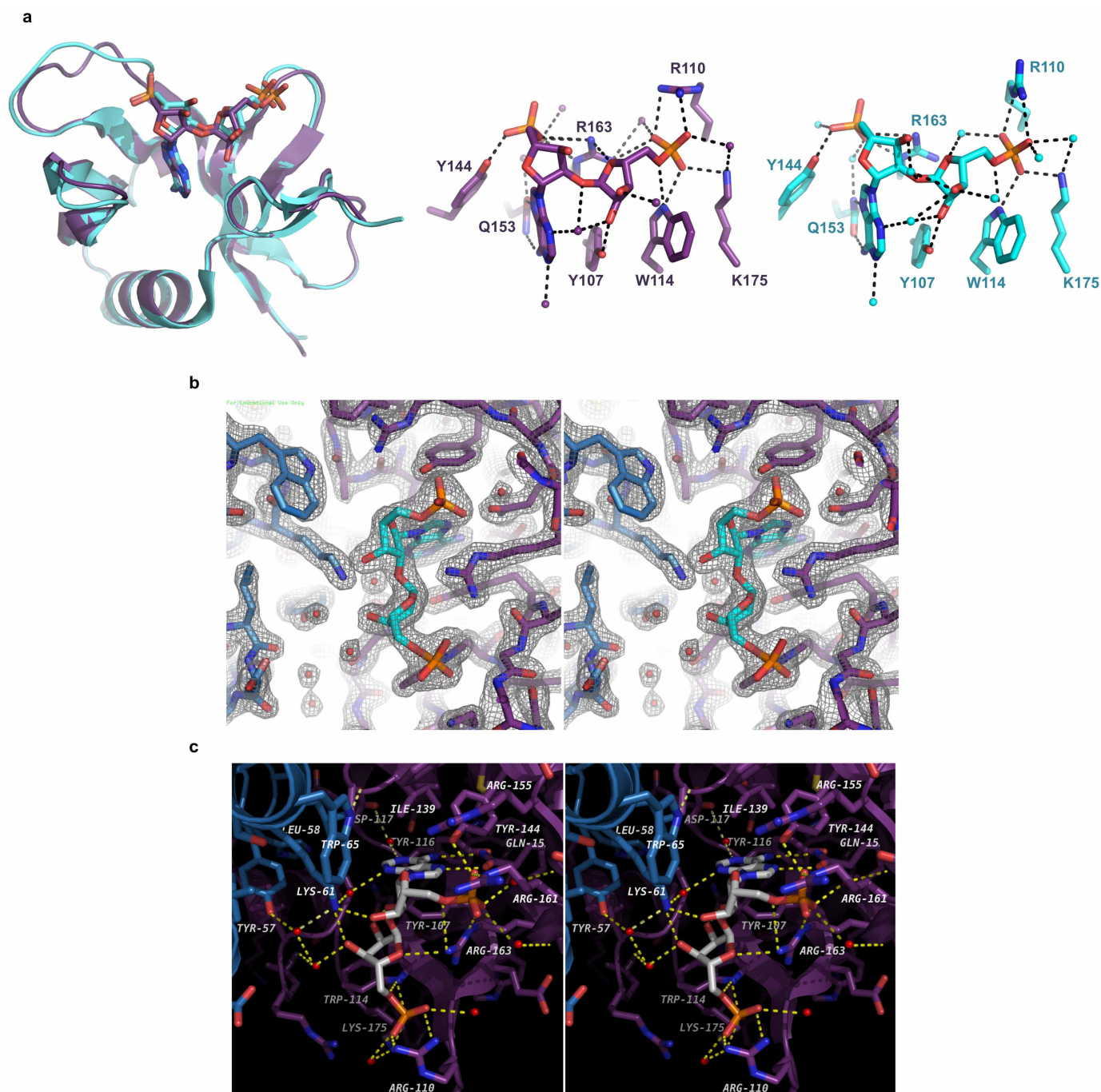
fragments: (left to right) wild type, Lys61Ala, and Lys61Asp. **c**, Limited proteolysis of RNF146(RING-WWE) and of a construct of RNF146 including the linker between the RING and WWE domains, and the WWE domain (RNF146(linker-WWE); residues 83–183). Both seem to result in the same product when treated with subtilisin. The RING-WWE construct is more resistant to subtilisin in the presence of ligand.



Extended Data Figure 4 | ^1H - ^{15}N HSQC-TROSY spectra of RNF146 reveal a conformational change in the RING domain after *iso*-ADPr binding.

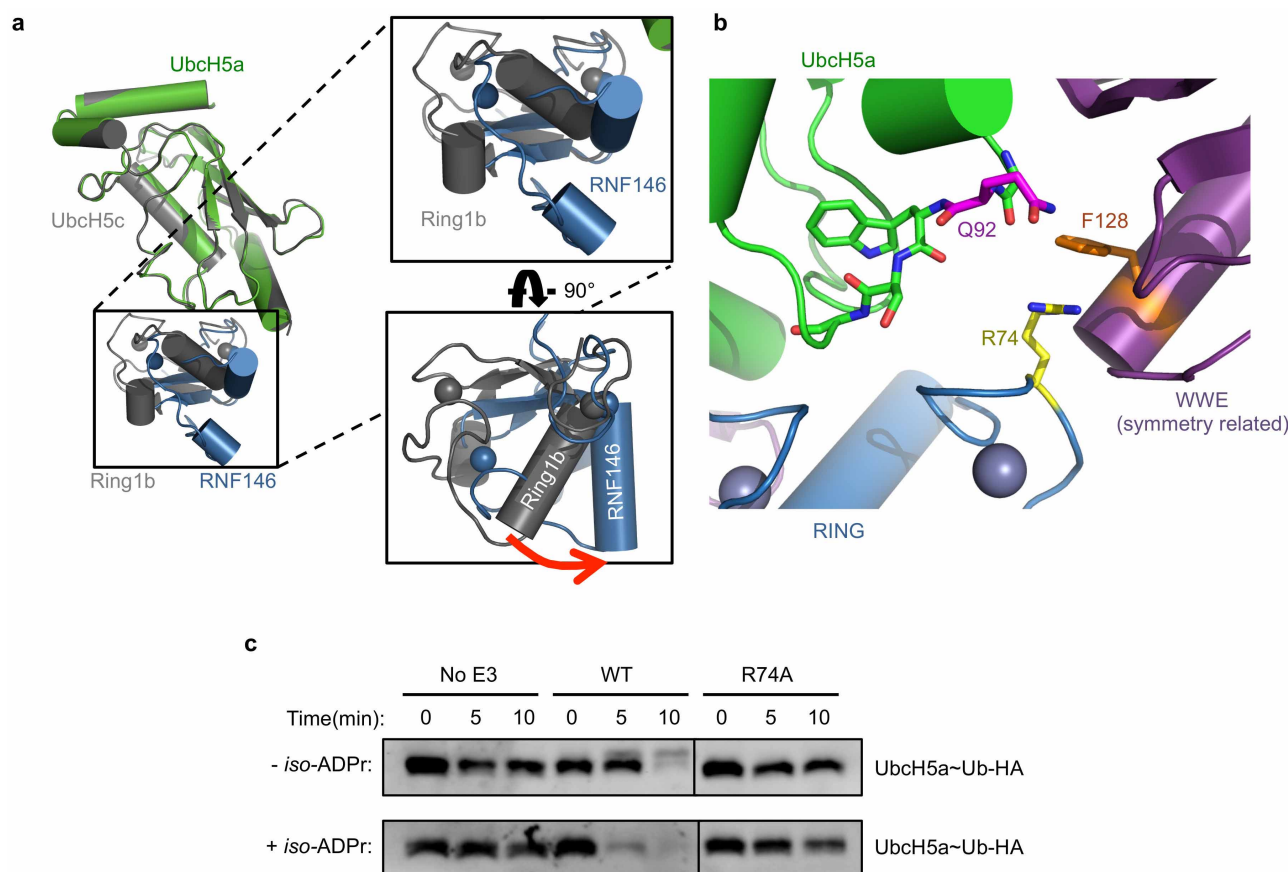
a, RNF146(RING-WWE) spectra in the absence (black) and presence (red) of saturating *iso*-ADPr concentrations show a marked change in most amide chemical environments. **b**, Overlay of the RNF146(RING-WWE) spectrum (black) with the spectrum of the RNF146(RING)-only domain (green) in the absence of *iso*-ADPr. Nearly all RING-only peaks overlay with a peak in the RING-WWE fragment spectrum, confirming that the RING-only domain adopts the same conformation as the RING domain in the larger fragment. **c**, Overlay of the liganded RING-WWE spectrum (red) with the isolated RING domain (green) shows very few corresponding peaks between the two spectra,

indicating environment changes of most RING domain peaks in the presence of *iso*-ADPr consistent with a conformational change. Notably, there are no changes in the spectrum of the RING-only construct when *iso*-ADPr is added under these conditions (data not shown). **d**, Close-up of **c** to illustrate that the RING domain samples a minor conformational state in liganded RNF146(RING-WWE). The minor peaks all correspond to RING peaks of unliganded RNF146 (black arrows). Therefore, the RING domain can still sample the non-activated conformation when saturated with ligand. Spectra were obtained with 200 μM protein and 300 μM *iso*-ADPr (saturating conditions).



Extended Data Figure 5 | Comparison of ligand binding in the RNF146(RING-WWE)-UbcH5a-*iso*-ADPr complex and in the WWE-only structure. **a**, Left, superposition of the WWE domain of the RNF146(RING-WWE)-UbcH5a-*iso*-ADPr complex (purple) and the previous *iso*-ADPr/WWE structure (cyan, PDB code 3V3L)¹³. Middle and right, WWE residues involved in binding *iso*-ADPr in the RNF146(RING-WWE)-UbcH5a-*iso*-ADPr complex (purple) (middle), and in the previous *iso*-ADPr/WWE structure (cyan) (right). Waters are shown as non-bonded spheres; hydrogen bonds are shown as dashed lines. Side-chain contacts between ligand and

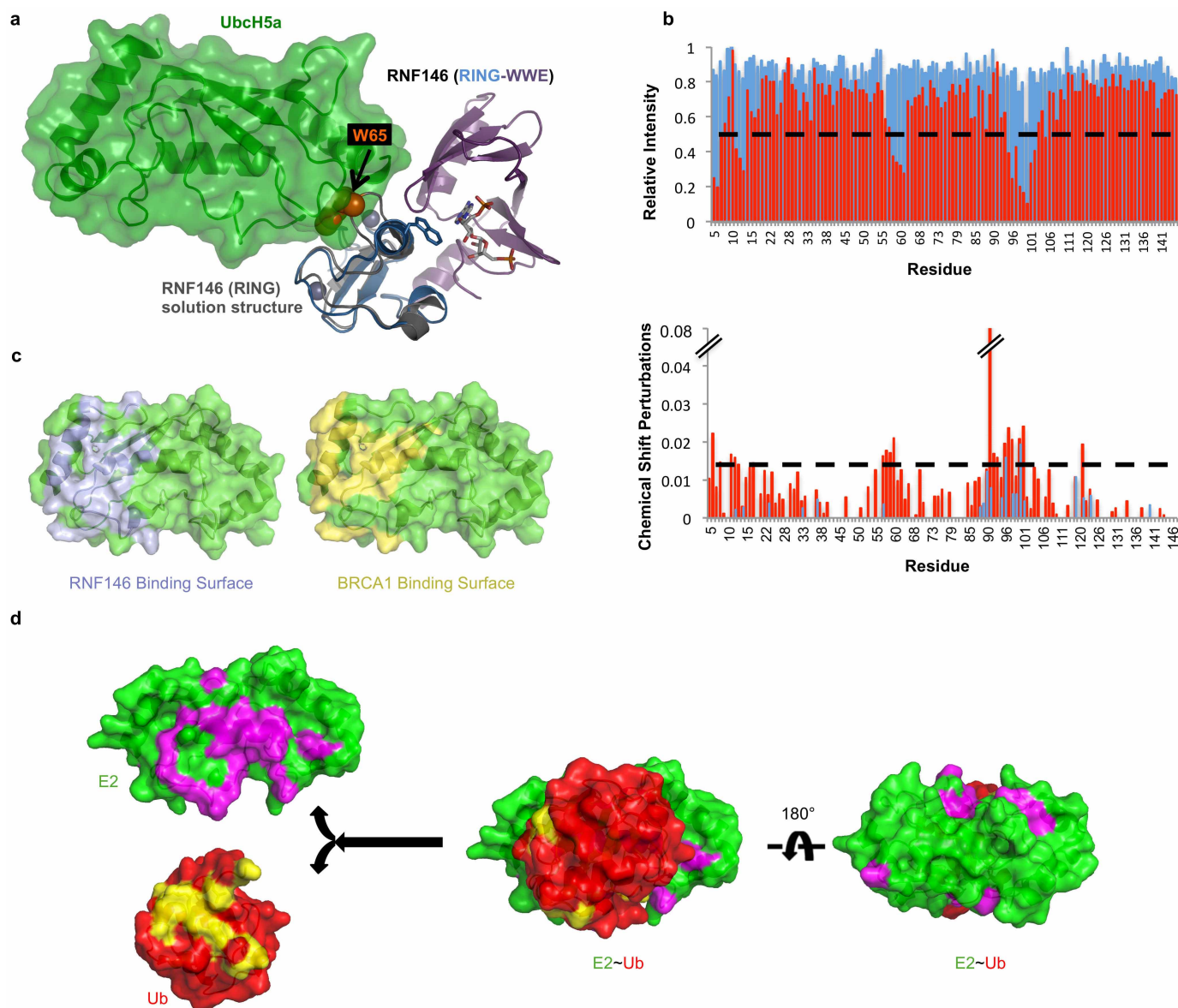
protein are maintained in both structures. **b**, Stereoview of the RNF146(RING-WWE)-UbcH5a-*iso*-ADPr complex ligand-binding site showing the $2F_o - F_c$ map (grey mesh) contoured at 1.5σ . The ligand and waters are well defined within the binding site. Waters are shown as red non-bonded spheres, *iso*-ADPr is shown in cyan, and the RING and WWE domains are coloured as in Fig. 2a. **c**, Stereoview of the *iso*-ADPr-binding site indicating residues within 4.5 Å of the ligand. Protein and ligand are represented as sticks, waters as red non-bonding spheres, and hydrogen bonds as dashed yellow lines. The RING and WWE domains and ligand are coloured as in Fig. 2a.



Extended Data Figure 6 | Rotation and crystal packing at the E2-E3 binding interface of the RNF146(RING-WWE)-UbcH5a-iso-ADPr complex.

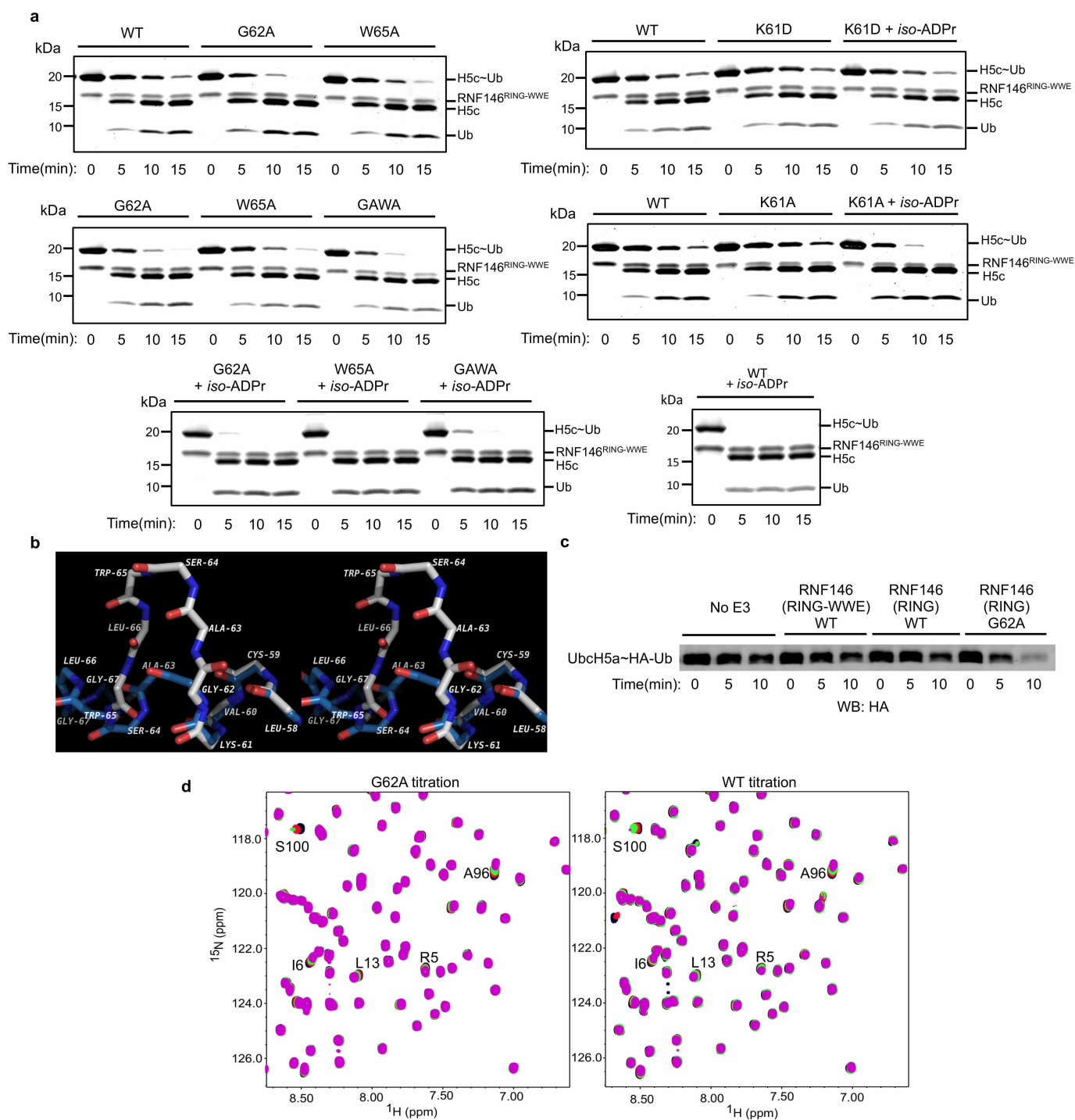
a, Superposition of the E2 in the RNF146(RING-WWE)-UbcH5a-iso-ADPr (coloured as in Fig. 2a) with a representative RING E3-E2 structure, the Bmi1-Ring1b-UbcH5c complex (grey) (PDB code 3RPG)¹⁸. The WWE domain is excluded for clarity. Boxes show close-up views of the RING domains revealing a rotation of the RING domain relative to the E2. Bottom right, RING domains rotated 90° to show the E2 binding surface of the E3s. The RING of the RNF146(RING-WWE)-UbcH5a-iso-ADPr structure is rotated relative to Ring1b-UbcH5c and other E3-E2 complexes^{15–20,23} (indicated by red arrow) when the E2s are aligned. **b**, Close-up view of the E2-E3 interface of the RNF146(RING-WWE)-UbcH5a-iso-ADPr complex shows that RING residue Arg74 (yellow) is too far (~7.7 Å) from the E2 Gln92 (magenta) carbonyl to

make the hydrogen bond observed in activated E3-E2~Ub structures^{17,21,22,24}. The side chain of Arg74 in the RING domain packs against Phe128 (orange), a WWE domain residue of a symmetry-related molecule in this crystal form. It is likely that crystal packing interferes with the formation of the 'allosteric' hydrogen bond. **c**, E2~Ub/lysine reactivity with RNF146(RING-WWE)(Arg74Ala) shows a dependence of RNF146 activity on the allosteric arginine^{17,21,22,24} with or without ligand. Because RNF146 activation requires Arg74, which does not make contacts with the E2 in our structure, and because RNF146 shows canonical E2 binding in solution (see Extended Data Fig. 7), we conclude that the orientation of RNF146 in the RNF146(RING-WWE)-UbcH5a-iso-ADPr complex is probably an unproductive E2-E3 association. The observed rotation is probably a crystallographic artefact.



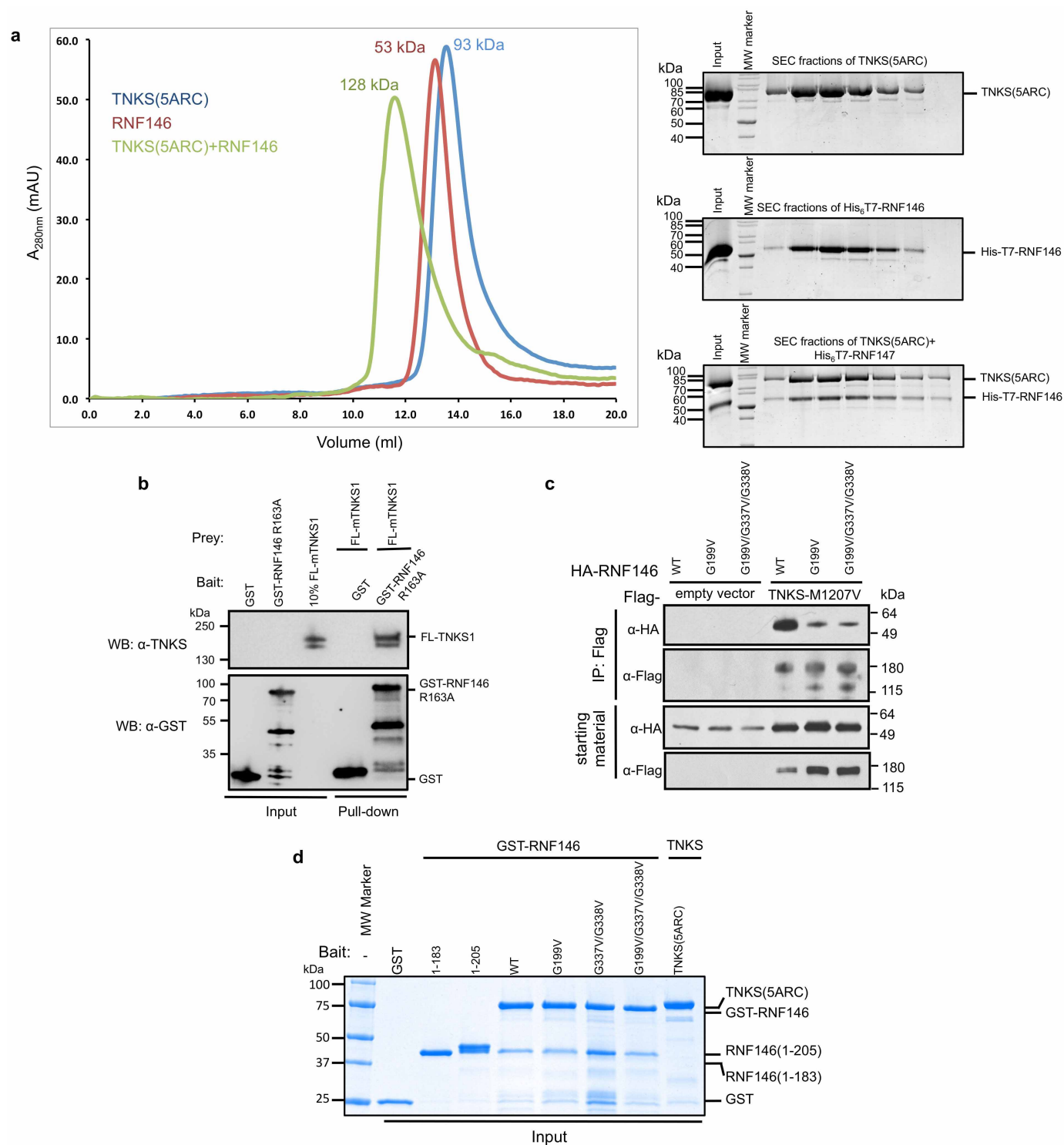
Extended Data Figure 7 | RNF146-*iso*-ADPr binding allows the RING domain to bind and activate a ubiquitin conjugating enzyme (E2). **a**, Left, superposition of the RING domain of unliganded RNF146 (PDB code 2D8T; grey, Trp 65 is shown as orange spheres) with the RNF146(RING-WWE)-UbcH5a-*iso*-ADPr complex (coloured as in Fig. 2a), shows a clash of Trp 65 with UbcH5a at the E2-E3 binding interface. This clash is observed when the RNF146(RING) structure (2D8T) is aligned with all other E2-E3 structures^{15–20,23}. **b**, Peak broadening (top; intensity relative to free E2) and CSPs (bottom) of ¹⁵N-UbcH5c(Ser22Arg/Cys85Ser) resonances (data are from the spectra shown in Fig. 3b). Histograms shown in blue compare the spectral properties of free E2 to E2 plus RNF146(RING-WWE); histograms shown in red compare free E2 to E2 plus RNF146(RING-WWE) and *iso*-ADPr. Dashed lines indicate one standard deviation from the mean value of the liganded (red) plots. Values below and above the dashed lines for the relative intensities and CSPs respectively are plotted on the E2 surface shown in **c** and Fig. 3b. **c**, Left, the RNF146(RING-WWE) binding surface inferred from data in **b** (light blue, on green E2), is compared with (right) the BRCA1-BARD1 binding surface on E2 (yellow, on green E2; residues 1–112 and residues 26–115,

respectively) previously inferred by an analogous experiment²⁵. When the NMR perturbations are mapped to the surface of UbcH5c, the revealed binding sites are very similar, and are consistent with previously reported binding surfaces for RING E3s on free ubiquitin conjugating enzymes^{15–20,23}. **d**, Chemical shift perturbations and broadening of resonances from ¹⁵N-E2-O-Ub conjugate (UbcH5c(Ser22Arg/Cys85Ser)-O-Ub) after RNF146(RING-WWE)-*iso*-ADPr binding (determined by the same method as shown in **b**, but with only 0.125 mol. equiv. E3 added to minimize hydrolysis of the E2-O-Ub oxyester during NMR data collection). Left, perturbed residues are mapped onto UbcH5b (magenta on green E2) and ubiquitin (yellow on red ubiquitin). Centre and right, perturbed residues mapped onto the structure of E2-O-Ub as it appears in the E3/E2-O-Ub complex of BIRC7-UbcH5b-Ub (PDB code 4AUQ; BIRC7 not shown for clarity)²⁴ show that the surfaces highlighted in the left panel are buried in the ‘closed’ state. The data show that RNF146 activates the E2~Ub conjugate by inducing the closed conformation^{17,21,22,24}. Because only the most perturbed residues are mapped to the E2~Ub surface, the E3 binding surface is not highlighted on the E2 in **d**.



Extended Data Figure 8 | Stabilizing helix 1 of RNF146 activates the RING domain. **a**, Complete images of gels shown in Fig. 3c (Oriole-stained) for Gly62Ala, Trp65Ala, Gly62Ala/Trp65Ala (GAWA), Lys61Ala, and Lys61Asp mutants of RNF146(RING-WWE) with or without *iso*-ADPr. Gly62Ala and GAWA mutants show reduced enhancement with *iso*-ADPr relative to wild type, probably owing to a clash of the Ala side chain with a turn in the WWE domain at position 62 (data not shown). **b**, Alignment of RNF146(RING) solution structure (PDB code 2D8T; white) and the crystal structure determined in this study (blue) shown in stereoview. Side chains are excluded for clarity; the backbone is represented by sticks. Comparison of the conformation of Gly62 in the two structures suggests a need for a small side

chain at position 62 to allow the structural transition from the inactive to active form of RNF146. **c**, Anti-HA western blot of the E2~Ub/lysine reactivity assay of RNF146(RING-WWE) compared with RNF146(RING) and RNF146(RING)(Gly62Ala) showing enhanced reactivity for the Gly to Ala mutation. **d**, Left, ^1H - ^{15}N HSQC-TROSY of ^{15}N -UbH5c(Ser22Arg/Cys85Ser) in the presence of 0.0 (black), 0.25 (red), 0.5 (green) and 1.0 (magenta) mol. equiv. of RNF146(RING)(Gly62Ala). Right, the same experiment performed with wild-type RNF146(RING). The most perturbed residues, indicated by letter and position (S100, etc.), show increased chemical shift perturbations for the RNF146(RING)(Gly62Ala) mutant.



Extended Data Figure 9 | RNF146 directly interacts with TNKS. **a**, Left, SEC profiles of untagged TNKS(5ARC) (blue), His₆T7-RNF146 (red), and a 1:1 mixture of these proteins (green). Numbers above the peaks indicate the average mass obtained by multi-angle static light scattering (MALS) for each peak. His₆T7-RNF146 and TNKS(5ARC) co-migrate as a single peak with an apparent mass of 128 kDa. Right, Coomassie-stained SDS-PAGE analysis of the SEC peaks in left panel show the presence of both proteins within the peak of the TNKS(5ARC)-His₆T7-RNF146 complex (bottom right). **b**, GST pull-down of partially purified full-length mouse tankyrase-1 (FL-mTNKS1) with GST-tagged RNF146(Arg163Ala) (PAR-binding deficient RNF146 mutant)¹¹.

Full-length mTNKS1 can be pulled down by GST-RNF146, but not GST. **c**, Co-immunoprecipitation of HA-RNF146 variants with transiently transfected flag-tagged TNKS(Met1207Val) (catalytically inactive mutant). The Met1207Val mutation prevents auto-PARylation of TNKS and therefore PAR-mediated interactions between RNF146 and TNKS³⁸. Under the experimental conditions, both the motif I mutant, Gly199Val, and the motifs I + IV mutant, Gly199Val/Gly337Val/Gly338Val, markedly reduce the RNF146-TNKS interaction. **d**, Coomassie-stained SDS-PAGE of proteins used in the GST pull-down assay shown in Fig. 4a (inputs). Samples were used in a 1:2 ratio (3 μM GST-RNF146 to 6.7 μM TNKS(5ARC)) for these GST pull-down experiments.

Extended Data Table 1 | Data collection, phasing and refinement statistics

	RNF146(RING-WWE)/UbcH5a/ <i>iso</i> -ADPr complex Zn²⁺ SAD
Data collection	
Space group	<i>P</i> 2 ₁ 2 ₁ 2
Cell dimensions	
<i>a</i> , <i>b</i> , <i>c</i> (Å)	133.67, 61.69, 94.35
α , β , γ (°)	90, 90, 90
Resolution (Å)	50.0 - 1.90 (1.96 - 1.90)*
<i>R</i> _{sym} (%)	9.2 (45.9)*
<i>I</i> / σ <i>I</i>	34.4 (2.5)*
Completeness (%)	95.6 (71.4)*
Redundancy	7.8 (4.6)*
Refinement	
Resolution (Å)	50.0 - 1.90
No. reflections (test set)	55779 (2994)
<i>R</i> _{work} / <i>R</i> _{free}	18.6 / 22.2
No. atoms	
Protein	4886
Ligand	72
Zn ²⁺	4
Water	371
B-factors	
Protein	16.3
Ligand	31.0
Zn ²⁺	31.5
Water	42.4
R.m.s deviations	
Bond lengths (Å)	0.009
Bond angles (°)	1.4

This diffraction data set was collected from a single crystal.

5% randomly selected reflections were used as a test set.

*Highest resolution shell is shown in parentheses.

The heat released during catalytic turnover enhances the diffusion of an enzyme

Clement Riedel¹, Ronen Gabizon¹, Christian A. M. Wilson^{1,2}, Kambiz Hamadani^{1†}, Konstantinos Tsekouras³, Susan Marqusee^{1,4}, Steve Presse^{3,5} & Carlos Bustamante^{1,4,6,7,8,9}

Recent studies have shown that the diffusivity of enzymes increases in a substrate-dependent manner during catalysis^{1,2}. Although this observation has been reported and characterized for several different systems^{3–10}, the precise origin of this phenomenon is unknown. Calorimetric methods are often used to determine enthalpies from enzyme-catalysed reactions and can therefore provide important insight into their reaction mechanisms^{11,12}. The ensemble averages involved in traditional bulk calorimetry cannot probe the transient effects that the energy exchanged in a reaction may have on the catalyst. Here we obtain single-molecule fluorescence correlation spectroscopy data and analyse them within the framework of a stochastic theory to demonstrate a mechanistic link between the enhanced diffusion of a single enzyme molecule and the heat released in the reaction. We propose that the heat released during catalysis generates an asymmetric pressure wave that results in a differential stress at the protein–solvent interface that transiently displaces the centre-of-mass of the enzyme (chemoacoustic effect). This novel perspective on how enzymes respond to the energy released during catalysis suggests a possible effect of the heat of reaction on the structural integrity and internal degrees of freedom of the enzyme.

Externally induced temperature spikes—through the use of laser pulses, for example—can have dramatic effects on enzyme catalysis and protein conformations¹³. Thus, pyramine and green fluorescent protein have been shown to blink with characteristic frequencies when excited with a laser. Some authors^{14,15} have related this blinking to local temperature and pH changes. Yet, no equivalent effect has been attributed to the heat exchanged in an enzyme-catalysed reaction despite the fact that some enzymes, like catalase, release enough heat to unfold a protein¹⁶.

Fluorescence correlation spectroscopy (FCS) results obtained in experiments similar to those presented here^{1,2} have demonstrated that the diffusion coefficient (D) of urease increases in the presence of its substrate¹. These authors explored various potential mechanisms that might account for this enhanced diffusion, including global temperature increase of the solution, charged product induced electrophoresis, and pH changes around the enzyme immediately following catalysis. Recently, the same group has shown that the diffusion coefficient of catalase, which mediates the conversion of hydrogen peroxide into water and oxygen, also increases in a substrate-dependent manner², ruling out charge or pH as a general explanation. They have also ruled out global or local temperature changes of the solution to explain the enhanced enzyme diffusion phenomenon¹. More recently, these authors have proposed that the enzyme diffusion coefficient increase arises from chemotactic behaviour in which the enzyme preferentially diffuses towards higher substrate gradients, although they provided no mechanism by which this may occur².

Here we performed a series of experiments and carried out a number of crucial controls to show that when enzymes catalyse reactions,

the heat released in the process is responsible for accelerating the protein's centre-of-mass, giving rise to the enhanced diffusion coefficient observed by FCS. We present a stochastic theory that predicts the linear dependence observed between the diffusion coefficient (measured by FCS) and the reaction rate V (measured by bulk enzymatic assays), and demonstrate that the coefficient of proportionality depends linearly on the enthalpy released by a single chemical reaction.

We studied four enzymes: catalase, urease, alkaline phosphatase and triose phosphate isomerase (TIM) (see Supplementary Information for more details on the enzymes). The rates of reaction catalysed by these enzymes and normalized by the enzyme concentration, were determined as a function of the substrate concentrations, $V([S])$, using bulk enzymatic assays (Extended Data Fig. 1). All enzymes followed Michaelis–Menten kinetics with k_{cat} and K_M values in agreement with those previously reported in the literature^{17–20} (Table 1).

FCS experiments of the fluorescently labelled enzymes were carried out in parallel to bulk assays to generate fluorescence intensity correlation functions, $G(\tau)$, as a function of substrate concentrations and initial velocities using previously described protocols for monitoring each enzymatic activity^{12,13}. These correlation functions were fit by the normal diffusion of a single species in a dilute solution to obtain the diffusion time (τ_D) and molecular occupancy of the labelled molecules within the confocal volume of observation²¹ (see Methods). Measurements were performed at a protein concentration of 1 nM for all enzymes. The FCS curves are shown in Fig. 1. The diffusion coefficient D can be calculated from the diffusion time τ_D using the radius, r , of the illuminated circular area crossed by the molecules according to: $D = r^2/4\tau_D$. The radius r was determined to be 500 nm by monitoring the diffusion of a free dye with a known diffusion coefficient through the illuminated area. Using the bulk assays to relate the substrate concentration to the enzyme's specific activity, we find a linear dependence between the normalized relative increase in diffusion coefficient, $(D - D_0)/D_0$, and the reaction rate for the enzyme systems (Fig. 2a–d).

Table 1 | Kinetic, calorimetric parameters of the enzymes and chemical reactions studied

	K_M (mM)	k_{cat} (s ^{−1}) × 10 ⁴	ΔH (kJ mol ^{−1})	α (m ²) × 10 ^{−16}
Catalase	62	5.8	−100	10 (9.6–11)
Urease	3	1.7	−59.6	4.5 (3.8–5.3)
Alkaline phosphatase	1.6	1.4	−43.5	16 (14–18)
Triose phosphate isomerase	1.8	1.3	−3.00	−2.5 (−6.1–1.1)

The Michaelis constant K_M and maximum turnover number k_{cat} are obtained from the fit to a Michaelis–Menten curve for the reaction rate as a function of substrate concentration (Extended Data Fig. 1). Values are in agreement with the ones from the supplier (Sigma) and found in the literature^{17–20}. ΔH is the enthalpy of the chemical reaction measured in bulk; values are reported in the literature for catalase³⁰, urease¹, triose phosphate isomerase²² and measured by isothermal titration calorimetry for alkaline phosphatase (see Extended Data Fig. 2). α is the coefficient of proportionality between the diffusion coefficient and the reaction rate, in parentheses is the 95% confidence interval.

¹California Institute for Quantitative Biosciences, QB3, University of California, Berkeley, California 94720, USA. ²Departamento de Bioquímica y Biología Molecular, Facultad de Ciencias Químicas y Farmacéuticas, Universidad de Chile, 1058 Santiago, Chile. ³Department of Physics, Indiana University–Purdue University Indianapolis (IUPUI), Indiana 46202, USA. ⁴Department of Molecular and Cell Biology, University of California, Berkeley, California 94720, USA. ⁵Department of Cellular and Integrative Physiology, Indiana University School of Medicine, Indiana 46202, USA. ⁶Jason L. Choy Laboratory of Single-Molecule Biophysics and Department of Physics, University of California, Berkeley, California 94720, USA. ⁷Department of Chemistry, University of California, Berkeley, California 94720, USA. ⁸Howard Hughes Medical Institute, University of California, Berkeley, California 94720, USA. ⁹Kavli Energy Nano Sciences Institute, University of California, Berkeley and Lawrence Berkeley National Laboratory, California 94720, USA. [†]Present address: Department of Chemistry and Biochemistry, California State University San Marcos, California 92078, USA.

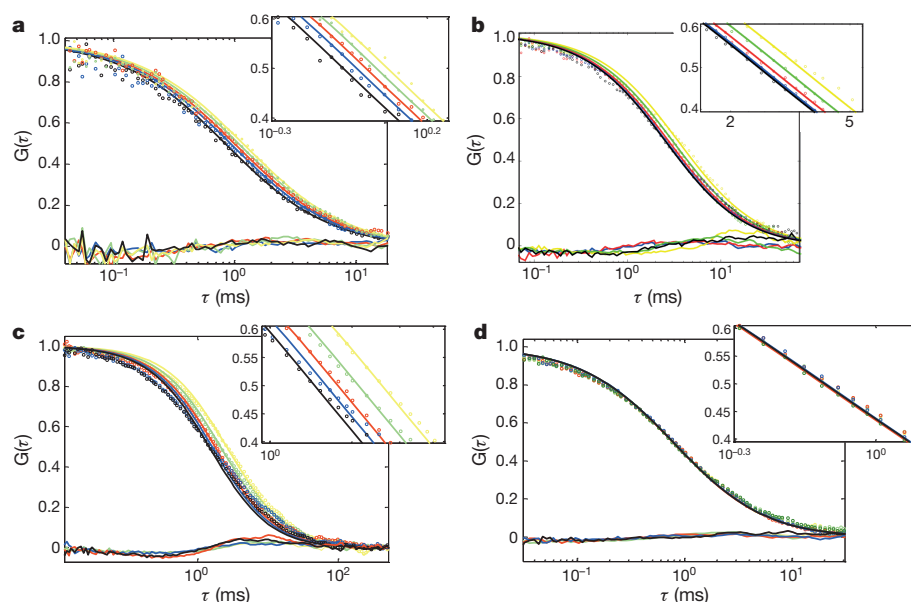


Figure 1 | FCS data and residuals of the fits. $G(\tau)$ is the normalized correlation function. Insets: zoom around the inflexion point. Lines are fits by a normal diffusion model with a single diffusive species. **a**, Catalase. Hydrogen peroxide concentration: yellow, 0 mM; green, 3 mM; red, 6 mM; blue, 12 mM; black, 25 mM. **b**, Urease. Urea concentration: yellow, 0 mM; green, 1 mM; red, 10 mM; blue, 100 mM; black, 1 M. **c**, Alkaline phosphatase. p-Nitrophenylphosphate concentration: yellow, 0 mM; green, 0.44 mM; red, 0.88 mM; blue, 1.74 mM; black, 2.64 mM. **d**, TIM. D-Glyceraldehyde 3-phosphate concentration: yellow, 0 mM; green, 0.15 mM; red, 0.31 mM; blue, 0.62 mM; black, 1.3 mM.

To confirm that the observed increase in diffusion coefficient is a result of the chemical reaction itself and not simply due to the binding or unbinding of the substrate, we performed control experiments with catalase using its non-competitive reversible inhibitor, sodium azide. FCS measurements were repeated for catalase in the presence of 5 mM sodium azide and a high concentration of substrate (25 mM hydrogen peroxide). Under these conditions, the enzyme is known to associate with and dissociate from the substrate without forming any product¹⁷. In the presence of inhibitor, we observe no increase in the diffusion coefficient relative to the measurements performed in the absence of substrate, indicating that the enhanced diffusion of catalase depends on the enzymatic reaction taking place and not just on the association or dissociation of the substrate to the enzyme.

To rule out other potential indirect effects on the apparent diffusion coefficient of the labelled diffusing enzymes (such as convective flows owing to oxygen bubbling, as in the case of catalase), we carried out experiments in which non-labelled, catalytically active catalase, in the presence of its substrate, was mixed with fluorescently labelled urease in the absence of its substrate. Increased diffusion of the labelled urease molecules was observed only under conditions when the concentration of hydrogen peroxide was greater than 100 mM and catalase concentration was above 5 nM. These results show that, below these concentrations, neither oxygen bubbling, nor global heating of the solution owing to the heat released by other molecules in the reaction, are responsible for the enhanced diffusion of enzymes observed in the presence of their substrate. Accordingly, all the results analysed here were obtained below these concentrations to ensure that only first-order, direct effects were being monitored (see Extended Data Fig. 3 for more details). Local heating of the solvent around the enzyme following a turnover event can also be eliminated because assuming the reaction heat is moved

into a 1 nm thick spherical shell surrounding catalase, which we assume to be a 4 nm radius spherical molecule, and using water's heat capacity, the local temperature of this shell would only increase by 0.15 K (see Supplementary Information).

Catalase, urease and alkaline phosphatase, which exhibit enhanced diffusion in the presence of their substrate (Fig. 2a–c), catalyse chemical reactions that are strongly exothermic (above 40 kJ mol^{-1}). Thus, we wondered whether the effect of the heat of the reaction on the enzyme itself could be responsible for the increased in diffusion coefficient observed upon catalysis. To test this idea we performed a negative control with triose phosphate isomerase (TIM). If the heat of the reaction is responsible for the increase in the diffusion coefficient observed, it follows that no such effect should be seen if the reaction is not exothermic or if it is only slightly so. TIM catalyses the reversible interconversion of the triose phosphate isomers dihydroxyacetone phosphate and D-glyceraldehyde 3-phosphate. The enthalpy of this reaction is small (-3 kJ mol^{-1} ; ref. 22). Consistent with the hypothesis formulated here, FCS experiments with TIM and diffusion coefficient analysis revealed no enhancement in the diffusion of the enzyme in the presence of its substrate throughout the range of substrate concentrations studied here (0 to 1.2 mM of D-glyceraldehyde 3-phosphate corresponding to reaction rates ranging from 0 to $\sim 6,000 \text{ s}^{-1}$) (Fig. 2d). Table 1 summarizes thermodynamic and kinetic parameters of each enzyme investigated.

On the basis of these observations, we propose that the increase in diffusion coefficient of the enzyme upon catalysis has its origin in an effect similar to that observed in photoacoustic spectroscopy. Here a vibrationally excited protein relaxes by dissipating its energy into the solvent through acoustic waves that are generated from the transient expansion and recompression of the protein immediately following excitation that can be detected by a microphone²³. Likewise, we propose that an enzyme expands, albeit asymmetrically, following the release of the heat of reaction (a process we call 'chemoacoustic effect'). The asymmetry is due to the location of the catalytic site with respect to the enzyme's centre-of-mass, as is the case in urease, catalase and alkaline phosphatase. The asymmetric pressure wave—following a catalytic event—should result in differential stress at the protein–solvent interface. The solvent's response is twofold: it dissipates energy through an acoustic wave and, more importantly, pushes back on the enzyme as dictated by Newton's third law, transiently displacing its centre-of-mass. In the Supplementary Material we compute an upper bound on the pressure exerted by the expanding protein on the solvent (about 500 pN nm^{-2}).

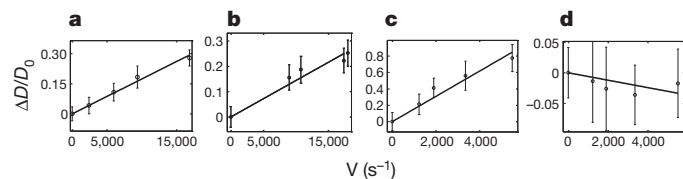


Figure 2 | Enhanced diffusion as a function of the reaction rate. **a–d**, Dimensionless change in diffusion coefficient ($\Delta D = (D - D_0)/D_0$) as a function of the reaction rate for catalase (**a**), urease (**b**), alkaline phosphatase (**c**) and triose phosphate isomerase (**d**). Error bars represent the standard deviation of 10 measurements.

To validate this interpretation, we performed experiments in which we directly excited the catalase haem group using a laser line at 402 nm (the Soret band of catalase is centred at 405 nm (ref. 24) and this transition possesses a fluorescence quantum yield $Q < 10^{-5}$)²⁵. In these experiments we sought to determine if the heat released upon radiationless de-excitation of the enzyme could also lead to an increased diffusion coefficient of the catalyst. Indeed, an increase of about 50% in the diffusion coefficient of the enzyme in the range of power between 0 and 1 mW was observed in these experiments (Extended Data Fig. 4), indicating that the local heat released by the haem in its transition from the excited to the ground state also generates a centre-of-mass motion of the enzyme (see Supplementary Discussion).

A stochastic model (detailed in the Supplementary Information and Extended Data Figs 6, 7 and 8) describes the enhanced diffusion upon catalysis in terms of the heat released by the chemical reaction. In this model, we assume that enzymes transiently diffuse more quickly—with diffusion coefficient D_1 —for some short period of time, δt , following a chemical reaction. Otherwise, the enzyme displays its diffusion coefficient in the absence of substrate, D_0 . The net diffusion coefficient in the presence of substrate, D , is therefore the ensemble average over both subpopulations with the probability of observing an enhanced diffusion proportional to V , the reaction rate. We then relate the enhanced diffusion coefficient, D_1 , to the amount of heat, Q , evolved by an enzymatic reaction. To do so, we assume that the kinetic energy of the enzyme's centre-of-mass immediately following a reaction is proportional to some fraction γ of Q . From this simple model, we obtain the following expression

$$D = D_0 + \frac{2\gamma Q}{3\zeta^2} mV \equiv D_0 + \alpha V$$

which shows a diffusion coefficient enhancement linear in V and Q , where m is the mass of the enzyme, $\delta t = m/\zeta$ is the relaxation timescale associated with the enzyme displacement following an enzymatic turnover, and ζ is an effective friction coefficient for the enzyme (see Supplementary Materials).

The parameter α describes the proportionality between the diffusion coefficient and the rate of the reaction. α itself depends on γ , the proportion of catalytic heat contributing to the enhanced translation of the centre-of-mass of the enzyme. As the protein's structure may dictate the precise mechanism by which enzymes dissipate catalytic heat, there is no simple correlation between the enhanced diffusion quantified by α and the enthalpy of the chemical reaction (see Table 1). For instance, the active sites of urease (two nickel atoms, see Extended Data Fig. 5a) are situated at the protein–solvent interface. By contrast, the haem sites of catalase (Extended Data Fig. 5b) are buried deep inside a highly conserved structure situated more than 20 Å from the nearest molecular surface²⁶. Accordingly, the heat released during catalysis may partition differently between the centre-of-mass and the internal degrees of freedom in these two proteins. In fact, time-resolved crystallography coupled with single-crystal microspectroscopy has been used to measure the correlation of electronic transitions with structural transitions during catalysis in catalase from *Proteus mirabilis*. This study indicates that the catalytic event is accompanied by small structural changes of the enzyme²⁷. It is possible that for catalase, part of the heat released during the chemical reaction is either absorbed by changes in the enzyme's conformation and/or dissipated through the protein's internal degrees of freedom, reducing its effect on the displacement of its centre-of-mass. Future simulations and experiments with different systems may shed light on the structural basis for energy dissipation and the partitioning of the heat of the reaction between the centre-of-mass and the internal degrees of freedom of the enzyme.

Here we considered catalytic reactions by globular proteins with freely diffusing substrates. However, many processive molecular machines (for example DNA polymerases, RNA polymerases, the ribosome, and so on) hydrolyse NTPs and generate forces to achieve directed motion along linear templates. It is interesting to speculate that some of these systems

may have evolved to absorb or indeed actively ‘channel’ the energy released by enzyme catalysis towards useful purposes (for example, enhanced diffusion, swimming, or even directed motion).

It is traditionally assumed that the heat released by a turnover event does not perturb an enzyme. Yet atomic force microscopy experiments have shown enhanced fluctuations of a cantilever tip placed atop a molecule of egg-white lysozyme in the presence of its substrate²⁸. The results presented here, in turn, indicate that the heat of the reaction can result in a centre-of-mass acceleration of the catalyst, suggesting a critical rethinking of how the reaction may affect the enzyme. For example, it is possible that the structural stability of catalase contributes to its high catalytic efficiency and observed turnover number, and that this parameter in many cases may be determined by the effect of the heat of the reaction on the structural integrity of the enzyme. In fact, studies of adenylate kinase have led to the suggestion of local unfolding of a protein domain near the active site by some unknown causing agent²⁹ during the enzymatic cycle—a phenomenon that has been termed “cracking”. Future fluorescence correlation spectroscopy–fluorescence resonance energy transfer experiments will probe how an enzyme's internal degrees of freedom and structural integrity are perturbed by the heat of the reaction.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 11 August; accepted 5 November 2014.

Published online 10 December 2014; corrected online 7 January 2015 (see full-text HTML version for details).

- Muddana, H. S., Sengupta, S., Mallouk, T. E., Sen, A. & Butler, P. J. Substrate catalysis enhances single-enzyme diffusion. *J. Am. Chem. Soc.* **132**, 2110–2111 (2010).
- Sengupta, S. *et al.* Enzyme molecules as nanomotors. *J. Am. Chem. Soc.* **135**, 1406–1414 (2013).
- Sánchez, S. & Pumera, M. Nanorobots: The ultimate wireless self-propelled sensing and actuating devices. *Chem. Asian J.* **4**, 1402–1410 (2009).
- Wang, J. Can man-made nanomachines compete with nature biomotors? *ACS Nano* **3**, 4–9 (2009).
- Howse, J. R. *et al.* Self-motile colloidal particles: from directed propulsion to random walk. *Phys. Rev. Lett.* **99**, 048102 (2007).
- Golestanian, R. Synthetic mechanochemical molecular swimmer. *Phys. Rev. Lett.* **105**, 018103 (2010).
- Kurtuldu, H., Guasto, J. S., Johnson, K. A. & Gollub, J. P. Enhancement of biomixing by swimming algal cells in two-dimensional films. *Proc. Natl Acad. Sci. USA* **108**, 10391–10395 (2011).
- Sakaue, T., Kapral, R. & Mikhailov, A. S. Nanoscale swimmers: hydrodynamic interactions and propulsion of molecular machines. *Eur. Phys. J. B* **75**, 381–387 (2010).
- Pavlick, R. A., Dey, K. K., Sirjoosingh, A., Benesi, A. & Sen, A. A catalytically driven organometallic molecular motor. *Nanoscale* **5**, 1301–1304 (2013).
- Yu, H., Jo, K., Kounovsky, K. L., Pablo, J. J. & Schwartz, D. C. Molecular propulsion: chemical sensing and chemotaxis of DNA driven by RNA polymerase. *J. Am. Chem. Soc.* **131**, 5722–5723 (2009).
- Todd, M. J. & Gomez, J. Enzyme kinetics determined using calorimetry: a general assay for enzyme activity? *Anal. Biochem.* **296**, 179–187 (2001).
- Williams, B. A. & Toone, E. J. Calorimetric evaluation of enzyme kinetic parameters. *J. Org. Chem.* **58**, 3507–3510 (1993).
- Steel, B. C., McKenzie, D. R., Bilek, M. M. M., Nosworthy, N. J. & dos Remedios, C. G. Nanosecond responses of proteins to ultra-high temperature pulses. *Biophys. J.* **91**, L66–L68 (2006).
- Wong, F. H. C., Banks, D. S., Abu-Arish, A. & Fradin, C. A molecular thermometer based on fluorescent protein blinking. *J. Am. Chem. Soc.* **129**, 10302–10303 (2007).
- Wong, F. H. C. & Fradin, C. Simultaneous pH and temperature measurements using pyranine as a molecular probe. *J. Fluoresc.* **21**, 299–312 (2011).
- Creighton, T. E. Protein folding. *Biochem. J.* **270**, 1–16 (1990).
- Switala, J. & Loewen, P. C. Diversity of properties among catalases. *Arch. Biochem. Biophys.* **401**, 145–154 (2002).
- Krajewska, B. & Ureases, I. Functional, catalytic and kinetic properties: a review. *J. Mol. Catal. B* **59**, 9–21 (2009).
- Go, M. K., Koudelka, A., Amyes, T. L. & Richard, J. P. Role of Lys-12 in catalysis by triosephosphate isomerase: a two-part substrate approach. *Biochemistry* **49**, 5377–5389 (2010).
- Sekiguchi, S., Hashida, Y., Yasukawa, K. & Inouye, K. Effects of amines and aminoalcohols on bovine intestine alkaline phosphatase activity. *Enzyme Microb. Technol.* **49**, 171–176 (2011).
- Elson, E. L. & Magde, D. Fluorescence correlation spectroscopy. I. Conceptual basis and theory. *Biopolymers* **13**, 1–27 (1974).

22. Goldberg, R. N., Tewari, Y. B. & Bhat, T. N. Thermodynamics of enzyme-catalyzed reactions—a database for quantitative biochemistry. *Bioinformatics* **20**, 2874–2877 (2004).
23. Peters, K. S. & Snyder, G. J. Time-resolved photoacoustic calorimetry: probing the energetics and dynamics of fast chemical and biochemical reactions. *Science* **241**, 1053–1057 (1988).
24. Haber, J., Maśalakiwicz, P., Rodakiewicz-Nowak, J. & Walde, P. Activity and spectroscopic properties of bovine liver catalase in sodium bis(2-ethylhexyl)sulfosuccinate/isooctane reverse micelles. *Eur. J. Biochem.* **217**, 567–573 (1993).
25. Sijja, L., Wei, M., Shasha, C., Gary, R. H. & Xie, X. S. Label-free imaging of heme proteins with two-photon excited photothermal lens microscopy. *Appl. Phys. Lett.* **96**, 113701 (2009).
26. Díaz, A., Loewen, P. C., Fita, I. & Carpena, X. Thirty years of heme catalases structural biology. *Arch. Biochem. Biophys.* **525**, 102–110 (2012).
27. Gouet, P. *et al.* Ferryl intermediates of catalase captured by time-resolved Weissenberg crystallography and UV-VIS spectroscopy. *Nature Struct. Mol. Biol.* **3**, 951–956 (1996).
28. Radmacher, M., Fritz, M., Hansma, H. G. & Hansma, P. K. Direct observation of enzyme activity with the atomic force microscope. *Science* **265**, 1577–1579 (1994).
29. Whitford, P. C., Miyashita, O., Levy, Y. & Onuchic, J. N. Conformational transitions of adenylate kinase: switching by cracking. *J. Mol. Biol.* **366**, 1661–1671 (2007).
30. Marzzacco, C. J. The enthalpy of decomposition of hydrogen peroxide: a general chemistry calorimetry experiment. *J. Chem. Educ.* **76**, 1517 (1999).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank A. Sen for discussions of his experiments.

We thank J. Kirsch, R. Golestanian, D. Leitner, J. Kirsdu, P. Geissler, P. Nelson and A. Szabo for their feedback on this work, J. Liphardt for the 405-nm laser line, P. Harbury for the TIM plasmid and B. Maguire for the TIM expression. This research was supported in part by NIH grants R01-GM0325543 (C.B.) and R01-GM05945 (S.M.), the US Department of Energy, Office of Basic Energy Sciences, Division of Materials Sciences and Engineering under contract no. DE-AC02-05CH11231 (C.B.), and the NSF grants MCB-1412259 (S.P.) and MCB-1122225 (S.M.). K.M.H. was supported in part by grant NIGMS, R01-GM65050. C.R. acknowledges the support of the Human Frontier Science Program. S.P. further acknowledges support from the Burroughs-Wellcome Fund.

Author Contributions C.R. performed fluorescence correlation spectroscopy (FCS) measurements, assisted in most experiments and was the primary writer of the manuscript. R.G. and C.A.M.W. ran most bulk biochemistry experiments; K.H. built the FCS setup and assisted in the FCS experiments; S.M. gave direction to the project. S.P. supervised the research and S.P. and K.T. developed the theory. Finally C.B. conceived the project, and supervised all of the research. All authors participated in the writing and editing of the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to S.P. (stevenpresse@gmail.com) or C.B. (carlosjbustamante@gmail.com).

METHODS

Labelling and purification. Catalase from bovine (*Bos taurus*) liver, urease from jack bean (*Canavalia ensiformis*) and alkaline phosphatase from bovine intestinal mucosa were purchased from Sigma Aldrich. Triose phosphate isomerase from *Saccharomyces cerevisiae* was expressed from a plasmid provided by P. Harbury. Catalase and urease were purified using a size exclusion column (Superdex 75 HiLoad 16/60 Prep Grade, Amersham Bioscience) and stored in sodium phosphate buffer at 0.1 M and pH 7.

Catalase and urease were labelled with Atto 647–maleimide (excitation/emission wavelengths: 647/665 nm, Molecular Probes). First, the protein was reacted with a tenfold excess of dithiothreitol for two hours at room temperature in order to reduce the cysteines. Dithiothreitol was removed using a desalting column (Micro Biospin 6 chromatography column, Biorad). The reaction with the dyes was then performed for four hours at room temperature and 12 h at 4 °C. Finally, the excess of dye was removed using a biospin column. Comparison of the absorbance at 280 nm and the absorbance at 647 nm indicated a labelling efficiency of one dye per molecule. Following the labelling, glycerol was added to 50% and the enzymes were stored at –20 °C.

TIM and alkaline phosphatase were labelled using Atto 647-NHS (*N*-hydroxy-succinimide). The proteins were dialysed against phosphate buffered saline (PBS), followed by reaction with dye in PBS + 20% DMSO for one hour at room temperature. Excess dye was removed by dialysis, and labelling efficiencies of 1–2 dyes per molecule were achieved. Labelled TIM was stored in MOPS 25 mM pH 6.5, 200 mM NaCl at –80 °C. Labelled alkaline phosphatase was stored in 5 mM Tris pH 7.9, 2.5 mM MgCl₂, 50 μM ZnCl₂ + 50% glycerol at 4 °C.

Bulk enzymatic assays. Bulk enzymatic assays were performed using a Jasco V 650 spectrophotometer at 25 °C.

For catalase, we used a 3-ml cuvette agitated with a magnet stir bar to reduce bubbling. The reaction rate *V* was calculated from the variation of absorbance (ΔA) of hydrogen peroxide (extinction coefficient $\epsilon = 43.5 \text{ M}^{-1} \text{ cm}^{-1}$ at 240 nm^{31,32}) and the enzyme concentration.

For urease, the reaction was stopped at different times by transferring 100 μl of the reaction mixture to a 1.5-ml microcentrifuge tube containing 500 μl of a 5% (w/v) phenol-nitroprusside solution. 500 μl of a 2% alkaline hypochlorite solution and 150 μl water were added, the mixture was then shaken, and the colour complex was allowed to develop for a minimum of 10 min at room temperature. Ammonia concentrations generated from the urease reaction, represented by the colour complex, were determined at 625 nm and compared to a standard curve made with NH₄Cl³³. Experiments were performed at a final concentration of enzyme $c_e = 10 \text{ nM}$.

Alkaline phosphatase assays were performed in diethanolamine 2 M pH 9.8, 1 mM MgCl₂ and 20 μM ZnCl₂, with 20 pM protein and different concentrations of p-nitrophenylphosphate (Sigma). The absorbance of nitrophenoxide at 405 nm was used to follow the reaction ($\epsilon = 18,000 \text{ M}^{-1} \text{ cm}^{-1}$).

TIM assays were performed in triethanolamine 100 mM pH 7.9, in the presence of 0.125 mM NADH, 4 units per ml of glycerol 3-phosphate dehydrogenase (Sigma), and 40 pM TIM. Different concentrations of DL-glyceraldehyde 3-phosphate (MPbio) were added and the change in absorbance of NADH at 339 nm ($\epsilon = 6,220 \text{ M}^{-1} \text{ cm}^{-1}$) was measured.

The activity of all labelled enzymes was compared to the non-labelled enzymes to ensure that the labelling did not affect the activity of the enzymes.

Isothermal titration calorimetry. The enthalpies of the reactions catalysed by TIM, catalase and urease in the conditions we used are published elsewhere^{1,22,30}. We used isothermal titration calorimetry to measure the enthalpy of the hydrolysis of p-nitrophenylphosphate catalysed by alkaline phosphatase in the buffer we used. Reactions were performed at 25 °C in diethanolamine 2 M pH 9.8, 1 mM MgCl₂

and 20 μM ZnCl₂, with 5 nM protein in the cell and 26.5 mM of p-nitrophenyl phosphate in the syringe. Aliquots of 2 μl (0.053 μmol) of substrate were added, and the released heat was measured for 15 min before adding more. The heat of the reaction was calculated by integrating the peaks and dividing by the amount of substrate added. Six additions were performed per experiment, and the experiment was carried out in triplicate. The heat of dilution of substrate alone without enzyme was negligible.

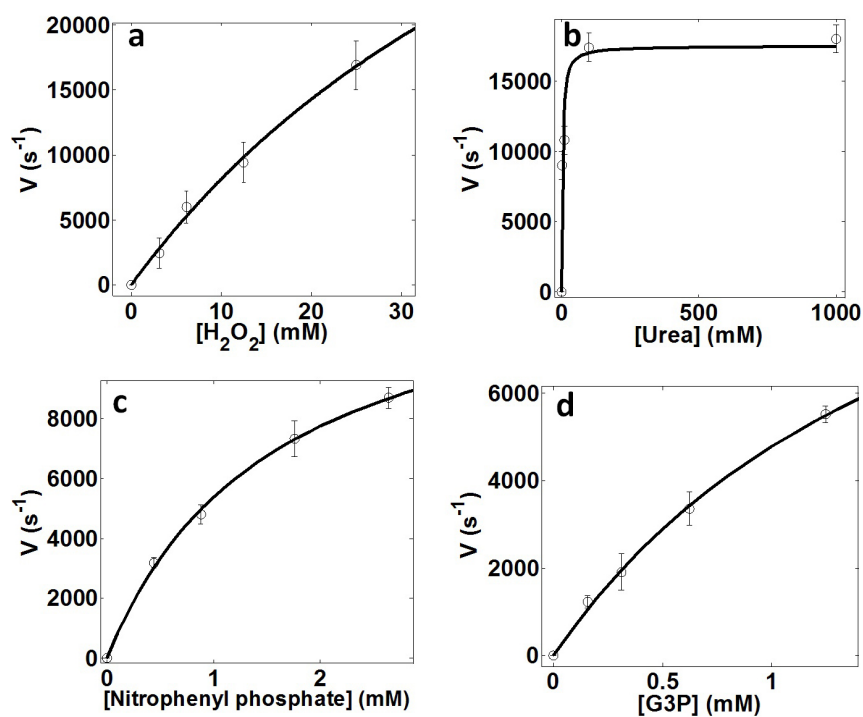
Fluorescence correlation spectroscopy. Fluorescence correlation spectroscopy measurements were performed with a purpose-built confocal microscope. Briefly, a 635 nm line from a high-stability diode LabLaser (Coherent Inc.) was coupled into a single-mode fibre (Thor Labs). The output from this fibre was collimated, reflected off of a polychroic mirror (z488/543/633rpc, Chroma Corp.) and under-filled ($\beta \sim 3$) into the back aperture of an infinity-corrected PSF-grade Uplan S apochromat 60× 1.2NA water-immersion objective (Olympus America) mounted atop a custom-machined microscope body. Typical incident laser powers were roughly 10–50 μW and were low enough to avoid dye blinking and/or bleaching effects. Emitted fluorescent photons were collected, focused onto a 100-μm pin-hole, collimated, split into two emission paths to minimize spurious correlations due to after-pulsing, and finally refocused onto the active areas of two avalanche photodiode (APD) detectors (Perkin Elmer Optoelectronics). The APD outputs were connected to the input channels of an ALV-5000 Multiple Tau Digital correlator. FCS correlation functions were acquired using ALV's commercial software suite. Experiments were performed at an enzyme concentration of 1 nM. At this nanomolar concentration, subunit dissociation can result in decreased enzyme activity. In order to control protein dissociation and have the same activity in both bulk and FCS experiments, enzymes were kept in ice and measurements were performed simultaneously, in the same substrate solution, at the same temperature (25 °C).

Data analysis. The ALV-5000 Multiple Tau Digital correlator software outputs the data points generating the fluorescence intensity correlation functions, $G(\tau)$, presented in Fig. 1 of the main manuscript. These data are fitted (using custom-built Matlab code) by the normal diffusion of a single species in a dilute solution³⁴:

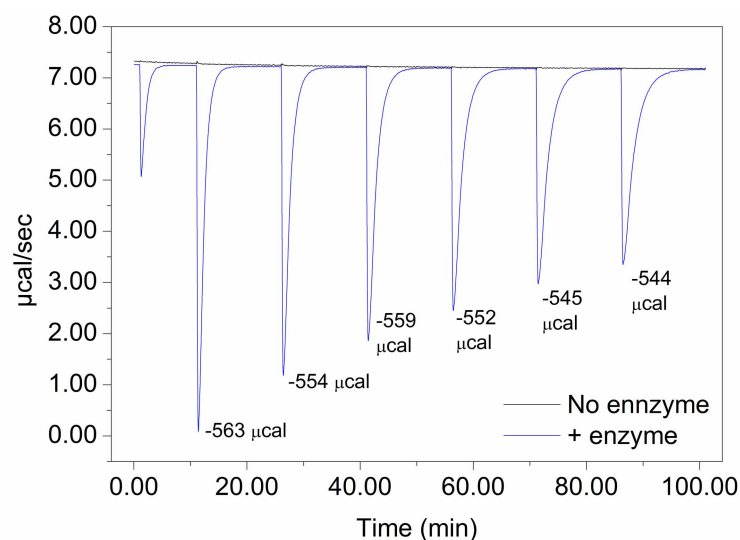
$$G(\tau) = \frac{1}{N} \left(1 + \frac{\tau}{\tau_D} \right)^{-1} \left(1 + \frac{\tau}{\omega^2 \tau_D} \right)^{-1/2}$$

where N is the average number of molecules in the confocal volume, τ_D is the characteristic diffusion time across the illuminated region and ω is the aspect ratio of the sampling volume (for our custom-built setup: $\omega = 5$). Therefore, the only two unknown parameters are N and τ_D . Although the enzyme's mobility shows small deviation from the free diffusion, as it can be seen in Fig. 1 the determination of τ_D is accurate. We are working at the single molecule level and N is randomly distributed between 1 and 3 for the four enzymes studied. Furthermore, we note that we are working in a regime with no second order effects - meaning that the activity of one enzyme does not affect the diffusion of another enzyme - and that the diffusion coefficient is therefore independent of N . Thus, τ_D is determined by the overall fit, and $D = r^2/4\tau_D$ where r is size of the beam waist perpendicular to the direction of the propagation of light³⁴.

31. Stern, K. G. On the absorption spectrum of catalase. *J. Biol. Chem.* **121**, 561–572 (1937).
32. Beers, R. F. & Sizer, I. W. A spectrophotometric method for measuring the breakdown of hydrogen peroxide by catalase. *J. Biol. Chem.* **195**, 133–140 (1952).
33. Smith, P. T., King, A. D. & Goodman, N. Isolation and characterization of urease from *Aspergillus niger*. *J. Gen. Microbiol.* **139**, 957–962 (1993).
34. Krichinsky, O. & Bonnet, G. Fluorescence correlation spectroscopy: the technique and its applications. *Rep. Prog. Phys.* **65**, 251 (2002).

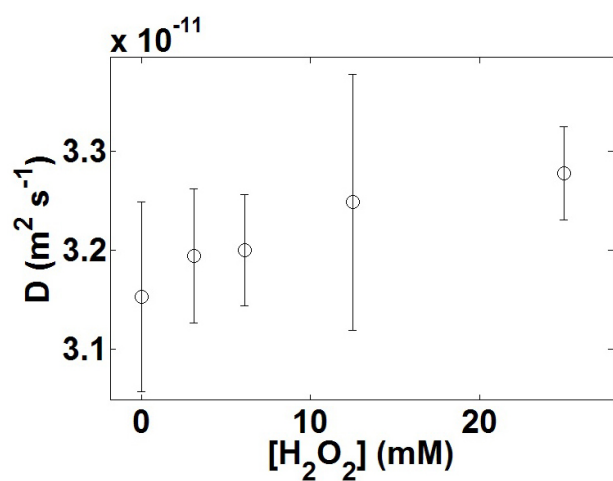


Extended Data Figure 1 | Reaction rate per enzyme molecule as a function of the concentration of substrate. a, Catalase; b, urease; c, alkaline phosphatase and d, triose phosphate isomerase. Lines are fit to a Michaelis–Menten curve. Error bars represent the standard deviation of 3 measurements.

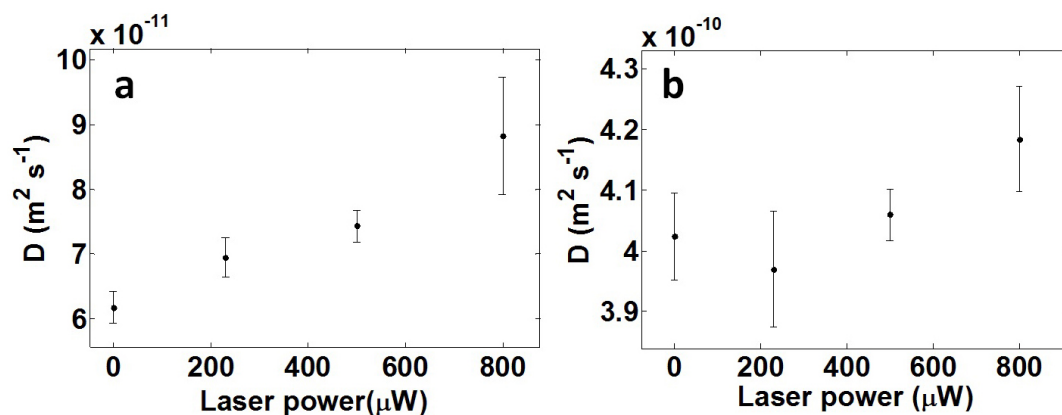


Extended Data Figure 2 | Isothermal titration calorimetry measurement of the heat of hydrolysis of p-nitrophenylphosphate by alkaline phosphatase. Each peak corresponds to the hydrolysis of 0.053 μmol of substrate. The

reaction rate slows as more substrate is added owing to the accumulation of phosphate product inhibiting the enzyme.

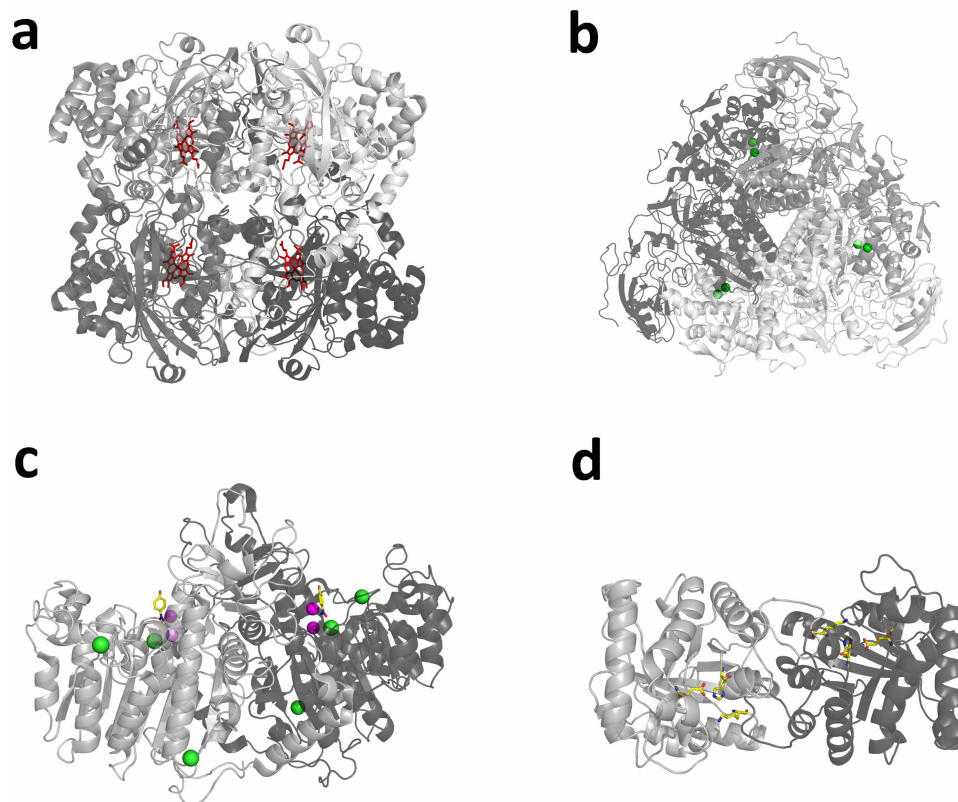


Extended Data Figure 3 | Diffusion coefficient of non-reactive urease in the presence of active, non-labelled, catalase (1 nM) for different concentrations of hydrogen peroxide. Even at the highest catalase activity no indirect effects (due to bubbling or global heating of the solution) appreciably increase the diffusion coefficient of freely diffusing urease. Error bars are computed from the standard deviation over 10 measurements.



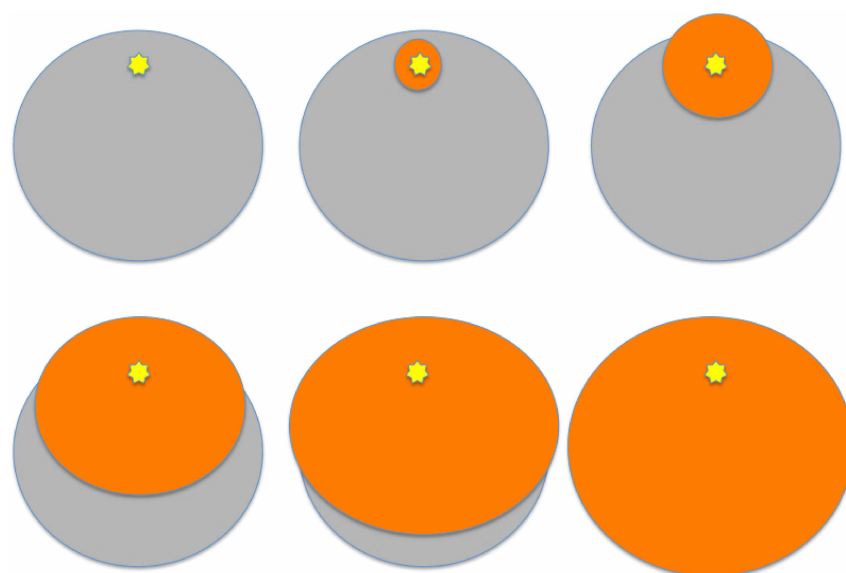
Extended Data Figure 4 | Haem excitation experiment. **a**, Diffusion coefficient of catalase as a function of the laser power at 402 nm. The heat released by the haem enhances the diffusion of the enzyme. **b**, Diffusion coefficient of free dyes in the presence of non-labelled heat-emitting

catalase. Neither the heat generated by the laser power nor the heat released by catalase significantly enhances the diffusion of the dyes. Error bars are computed from the standard deviation over 10 measurements.

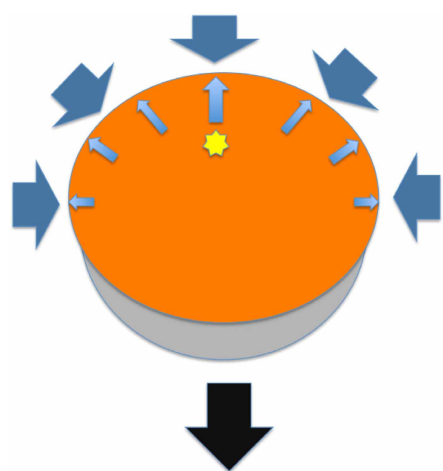


Extended Data Figure 5 | Structures of the enzymes studied with active site components highlighted. Different shades of grey indicate different monomers. **a**, Catalase (PDB 3NWL). Haem groups are shown as red sticks. **b**, Urease (PDB 4GY7). Active site nickel ions are shown as green spheres.

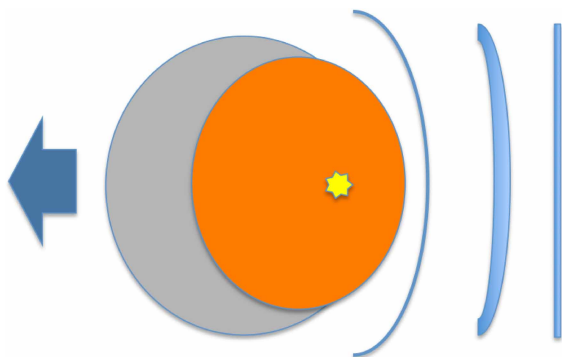
c, Alkaline phosphatase (PDB 4KJG). Magnesium ions are shown in green; zinc ions are shown in magenta; 4-nitrophenol (hydrolysis product bound at the active site) is shown as yellow sticks. **d**, Triose phosphate isomerase (8TIM). Active site residues Lys 12, His 95 and Glu 165 are shown as yellow sticks.



Extended Data Figure 6 | Protein deformation progression. Spread of a radial deformation wave (red) emanating from the catalytic site (yellow star). The protein (grey) has an approximate radius of 4 nm and the deformation wave speed is an estimated 3 nm ps^{-1} .



Extended Data Figure 7 | Centre of mass translation. The protein exerts forces on the solvent (small blue arrows). The solvent restoring forces (thick blue arrows) give rise to a protein centre of mass translation (black arrow).



Extended Data Figure 8 | Acoustic wave due to deformation. Here we assume the protein deforms and this generates acoustic waves in the solvent. If these acoustic waves were reflected back on the protein, they would give rise to a centre of mass translation denoted by the blue arrow.

CAREERS

DECISIONS Women are more likely to lose interest in research careers **p.233**

GRANTS New NIH requirements put past papers in context **p.233**

NATUREJOBS For the latest career listings and advice www.naturejobs.com



ADAPTED FROM PAZLANISME/SHUTTERSTOCK

OUTREACH

Speak up for science

Whether publishing contentious findings or defending evidence, the right tone is essential.

BY VIRGINIA GEWIN

David Robert Grimes, a postdoctoral researcher at the University of Oxford, UK, is an adamant defender of science, but the blog post he wrote in August caused quite a stir, even for him. Troubled by an upcoming vote by Dublin City Council on whether to stop fluoridating the city's water supply, in his post he implored councillors to heed evidence that fluoridated water helps to prevent and slow tooth decay. His contention that claims to the contrary are inflammatory, invalid and dishonest prompted critics around the world to call for his resignation, he says.

Grimes's experience may be extreme, but it serves as a reminder that publicly presenting controversial findings or opinions — even

if they are factually sound — can leave one open to hostility. When it comes to speaking up, the uncomfortable reality is that sometimes scientists, especially those at trainee and junior levels, face a difficult choice: courting controversy can exact a career toll, but staying mute can lead to harmful policy decisions uninformed by science.

Early-career researchers who want to speak out must learn how best to communicate science and how to respond productively to criticism, experts say. Yet many young researchers engage the public armed with little communications or media training (see *Nature* **468**, 465–467; 2010), and when wading into combustible topics such as climate change, evolution and public health, they quickly learn that science often runs counter to strongly

held ideology. Industrial factions with vested interests may mount a campaign to debunk the science — and sometimes even scientific colleagues can come out swinging if they hold opposing views or feel that an outspoken colleague is crassly seeking the spotlight.

Grimes knows all about the negative consequences of putting one's head above the parapet. He has written evidence-backed opinion articles on provocative topics such as climate change and abortion rights, and in return he has received angry e-mails and several threats of bodily harm. On a couple of occasions, he had to take out a restraining order against harassers. For his efforts in standing up for science, in October 2014 he was one of two recipients of the John Maddox Prize, which is an initiative of *Nature* and the philanthropic ►

► Kohn Foundation that recognizes efforts to promote evidence on matters of public interest.

Most researchers will not be as willing as Grimes to put themselves in the firing line, but at some point many scientists will have to present findings that are deemed controversial, or they might feel the need to stand up for the evidence, as Grimes did. At this point they need an adviser who can provide support when they engage with the media — and there are also several approaches that can be used to head off hostility before it has time to gain strength.

HEIGHT OF CONDESCENSION

When Grimes first began to write opinion pieces three years ago, he admits that he was slightly naive. “I thought the public suffered from an information deficit and all I needed to do was simply inform them what the science says on matters of policy,” he says. He quickly learned that expecting people to make decisions based solely on evidence is presumptuous at best. Worse, scientists who take that approach may sound condescending. “Scientists have a

“There is blood, sweat and tears, but ultimately we reward people that make science stronger.”

highly technical voice that can come across as talking down to the audience, even if they don’t mean to,” says Kathleen Hall Jamieson, director of the Annenberg Public Policy Center at the University of Pennsylvania in Philadelphia. Likewise, if scientists seem self-interested or partisan, they risk undermining their credibility.

Jamieson decided to test whether scientists could improve the acceptance of their message. In September 2014, she published a model science-communication strategy¹ that she tested using the particularly divisive topic of Arctic sea-ice trends. The model assumes that showing respect for an audience — such as avoiding technical language or jargon that could be perceived as ‘talking down’ — is key to making a strong case for the science. She found that researchers can boost acceptance of their message when they rely on highly regarded, apolitical sources of information and visually present statistics that invite the audience to draw their own conclusions (see ‘Tough sell’).

Boris Worm, a marine ecologist at Dalhousie University in Halifax, Canada, agrees that evidence is key to maintaining one’s voice even when under attack. In 2006, he published a controversial paper² that suggested that declines in global fisheries were impairing the ocean’s ability to provide food and maintain water quality. Established fisheries biologists began to criticize his work.

He remained impassive. “When paradigms are challenged, it can get really heated emotionally, but while it can come across as a personal attack, it’s not,” he says. “It’s about the results.” If anything, severe scrutiny can be seen as a sign that the pursuit really matters, he argues.

But Worm adds that it is crucial to embrace criticism if it is valid and not simply mudslinging. “Don’t try to run from the criticism; instead, directly engage it,” he says. In his case, that meant writing a paper with his critics³. He is as proud of that paper as he is of the initial work. “There is blood, sweat and tears, but ultimately we reward people that make science stronger and better and give them our respect,” he says.

Nancy Baron is director of science outreach at COMPASS, an organization based in Portland, Oregon, that provides media training to scientists. She advises her clients to consider sharing bits of pending publications with known critics to get their reactions beforehand — that way they can prepare for potentially acerbic push back, or possibly circumvent it. “This acknowledgement is often seen as a sign of respect, which can take some vehemence away and lead to a much more cordial discussion,” she says.

EMBRACE THE CRITICS

Evolutionary biologist Kevin Laland at the University of St Andrews, UK, took this tack. He had grown weary of fending off sceptics who, he felt, were misrepresenting his theory on how organisms can change their environment to produce evolutionary and ecological consequences. So he invited them to co-write a paper with him. After 26 rounds of editing, they produced an article that laid out the nature of the dispute, including where the two camps agreed and where they did not⁴. “I think this approach is far more helpful than attacking one another and throwing smoke,” he says.

In retrospect, Martin Krkosek, now a population ecologist at the University of Toronto in Canada, wishes that he had taken a similar approach and had let colleagues know of his provocative findings as a doctoral student in 2005. Instead, he floundered on the release of a paper⁵ that linked aquaculture operations to sea-lice infections in wild salmon. Colleagues in his field found themselves in the middle

of a media storm that they had no idea was brewing. “I worry, now, that when I put out a grant proposal it lands on the desk of someone whom I made life hard for,” he says. “Letting them know this paper was coming would have been a nice olive branch.”

The experience was a learning process for him, and he came to realize that it is important to address all manner of negative feedback. When the paper came out, he and his co-authors chose to respond only to criticisms raised formally in the scientific literature, not in the popular media. “We just didn’t respond to all the stuff that came out in the press — and we realized that our lack of follow-up may have diminished the work’s impact,” he says, because for one thing, it was easier for industry voices in the press to question whether aquaculture was truly to blame.

When it came to releasing two follow-up papers, Krkosek and his co-authors were better prepared. “The first thing we did was create time to deal with it,” he says of the media onslaught. They engaged in lengthy e-mail correspondences with critics, responded to media requests for interviews and developed a web page where they could post answers to frequently asked questions. The website, he says, was helpful because it gave them a venue to explain results in more detail and to supplement their findings with new analyses.

Ultimately, the evidence in the three papers, corroborated and further explained in press interviews, on the web page and in individual correspondence, helped to compel changes in management and policies by the British Columbia government that benefited wild salmon: changes that included a moratorium on further aquaculture development.

“I had a great sense of satisfaction that we had a real impact on the ground,” says Krkosek. And that is what Baron counsels young scientists to do — to take the long view. “Dealing with backlash is no fun, but in the end the result is often worth it,” she says. And there can be

TOUGH SELL

How to speak out on controversial topics

If you want to get your message across, it pays to know the most effective ways to engage. Here are a few tips from researchers, including communication expert Kathleen Hall Jamieson of the University of Pennsylvania in Philadelphia.

- Leverage your scientific persona. Viewers and listeners will believe that a scientific source is credible if they also believe that the scientist is respectful and is not trying to ‘sell’ or persuade them.
- Walk the audience through the logic behind your conclusions.
- Show the audience the clear and

irrefutable data trends. Include all data — omissions allow critics to call your motives into question.

- Translate the science into familiar concepts for all audiences.
- Anticipate tough questions and be prepared with answers.
- Set aside time to respond to media and critics.
- Create a website with frequently asked questions and responses.
- Embrace valid criticism and respond to it unemotionally.
- Consider reaching out to the opposition. **V.G.**



Meteorologist J. Marshall Shepherd, host of *Weather Geeks*, is comfortable advocating for science.

unforeseen benefits: Krkosek says that he has become more adept at separating his emotional and intellectual responses — a skill that has been useful in both his career and his personal life.

Kathie Dello, associate director of the Oregon Climate Change Research Institute in Corvallis, has also learned how to deal with criticism. She gives public talks that highlight the regional impacts of climate change, and she has had some harrowing in-person and e-mail encounters with the public and researchers, in which her understanding of the science behind her assertions has been questioned. After a negative encounter, she allows herself only a short amount of time to process it and to salvage any lessons from the experience. (She has learned, for example, to request a moderator at public forums.) Then she moves on. She has also learned that a game of dodgeball can be cathartic.

VALUED OPINIONS

There is little doubt that the greater a scientist's standing, the greater the impact their words will have — and, conversely, the lower a scientist's standing, the more potentially devastating any fallout from their words will be for their career. Perhaps not surprisingly, many scientists wait until they are in senior positions to use their voices.

J. Marshall Shepherd, a meteorologist at the University of Georgia in Athens, had published dozens of papers and had received a presidential award of excellence, but becoming president of the American Meteorological Society gave him a platform to amplify his voice. He now writes opinion pieces, is active on social media and hosts a television show called *Weather Geeks*. He advises younger scientists to do two things when deciding whether to speak up for science — establish

scientific credibility and carefully evaluate whether they have the unshakeable temperament that is necessary for times when they may be challenged.

Wildlife biologist Chris Darimont at the University of Victoria in Canada frequently speaks out on trophy hunting and other controversial issues that he has studied, and has engaged in live radio and television debates with trophy-hunting advocates. But before he secured his tenure-track position, he worried about the backlash he could receive for such public speaking, and the long fingerprints it would leave on the web. “When people are assessing you for a job, it's not just your CV they are reviewing, it's any controversy online as well,” he says. He eventually concluded that efforts to engage constructively in the public sphere could shine through, too. “I accepted that I would take the good with the bad,” he says. He also finds it productive to engage with policy-makers. “Thinking some decision makers are going to find your paper in a journal, understand it and use it to make evidence-based decisions borders on absurd,” he says.

Ultimately, scientists who speak out create room at the table for evidence, say those who have found their voice. “If we aren't there speaking on the science,” says Shepherd, “people skilled in messaging, such as attorneys and lawyers, will fill the gaps.” ■

Virginia Gewin is a freelance writer based in Portland, Oregon.

1. Hall Jamieson, K. & Hardy, B. W. *Proc. Natl Acad. Sci. USA* **111** (Suppl. 4), 13598–13605 (2014).
2. Worm, B. *et al. Science* **314**, 787–790 (2006).
3. Worm, B. *et al. Science* **325**, 578–585 (2009).
4. Scott-Phillips, T. C. *Evolution* **68**, 1231–1243 (2014).
5. Krkosek, M., Lewis, M. A. & Volpe, J. P. *Proc. Biol. Sci.* **272**, 689–696 (2005).

CAREER DECISIONS

Of academic interest

A survey of 1,500 recent US biomedical PhD graduates found that white and Asian men, who are well represented in academic institutions, are most likely to show interest in becoming academic researchers (K. D. Gibbs *et al. PLoS ONE* **9**, e114736; 2014). Although scientists of all ethnicities reported losing interest in faculty careers as their doctoral studies continued, women's loss of interest was more pronounced, particularly for underrepresented minorities. The trend persisted after controlling for factors such as publication record and a sense of belonging. Understanding why career interests differ is crucial for increasing diversity, says study co-author Kimberly Griffin of the University of Maryland, College Park.

GRANTS

Paint me a picture

Applicants for US National Institutes of Health (NIH) grants will now be asked to contextualize past work, which could help early-career researchers. A revised ‘biosketch’ section instructs scientists to explain how their experience will allow them to accomplish the proposal's goals. Sally Rockey, the NIH's deputy director for extramural research, has said that the change puts focus on an applicant's skills and accomplishments, not just their publications. The format is being rolled out in January and will become mandatory for most NIH grants on 25 May. Online tools such as SciENcv have been updated to help applicants to create biosketches that can be used across various government agencies.

HIGHER EDUCATION

UK science numbers up

Between 2003 and 2012, UK student enrolment in the biological and physical sciences rose by more than 30% and overall university enrolment rose by more than 6%, according to a report by the country's Higher Education Statistics Agency. But across all fields, the number of postgraduate researchers increased by only 0.1% for 2012–13, the smallest increase since 2004. In the decade since 2003, enrolment at UK institutions fell sharply for students from the Middle East (–124%), Asia (–63%), Africa (–36%) and North America (–31%), but rose for students from Australasia (22%) and South America (21%).



Meteorologist J. Marshall Shepherd, host of *Weather Geeks*, is comfortable advocating for science.

unforeseen benefits: Krkosek says that he has become more adept at separating his emotional and intellectual responses — a skill that has been useful in both his career and his personal life.

Kathie Dello, associate director of the Oregon Climate Change Research Institute in Corvallis, has also learned how to deal with criticism. She gives public talks that highlight the regional impacts of climate change, and she has had some harrowing in-person and e-mail encounters with the public and researchers, in which her understanding of the science behind her assertions has been questioned. After a negative encounter, she allows herself only a short amount of time to process it and to salvage any lessons from the experience. (She has learned, for example, to request a moderator at public forums.) Then she moves on. She has also learned that a game of dodgeball can be cathartic.

VALUED OPINIONS

There is little doubt that the greater a scientist's standing, the greater the impact their words will have — and, conversely, the lower a scientist's standing, the more potentially devastating any fallout from their words will be for their career. Perhaps not surprisingly, many scientists wait until they are in senior positions to use their voices.

J. Marshall Shepherd, a meteorologist at the University of Georgia in Athens, had published dozens of papers and had received a presidential award of excellence, but becoming president of the American Meteorological Society gave him a platform to amplify his voice. He now writes opinion pieces, is active on social media and hosts a television show called *Weather Geeks*. He advises younger scientists to do two things when deciding whether to speak up for science — establish

scientific credibility and carefully evaluate whether they have the unshakeable temperament that is necessary for times when they may be challenged.

Wildlife biologist Chris Darimont at the University of Victoria in Canada frequently speaks out on trophy hunting and other controversial issues that he has studied, and has engaged in live radio and television debates with trophy-hunting advocates. But before he secured his tenure-track position, he worried about the backlash he could receive for such public speaking, and the long fingerprints it would leave on the web. “When people are assessing you for a job, it's not just your CV they are reviewing, it's any controversy online as well,” he says. He eventually concluded that efforts to engage constructively in the public sphere could shine through, too. “I accepted that I would take the good with the bad,” he says. He also finds it productive to engage with policy-makers. “Thinking some decision makers are going to find your paper in a journal, understand it and use it to make evidence-based decisions borders on absurd,” he says.

Ultimately, scientists who speak out create room at the table for evidence, say those who have found their voice. “If we aren't there speaking on the science,” says Shepherd, “people skilled in messaging, such as attorneys and lawyers, will fill the gaps.” ■

Virginia Gewin is a freelance writer based in Portland, Oregon.

1. Hall Jamieson, K. & Hardy, B. W. *Proc. Natl Acad. Sci. USA* **111** (Suppl. 4), 13598–13605 (2014).
2. Worm, B. *et al. Science* **314**, 787–790 (2006).
3. Worm, B. *et al. Science* **325**, 578–585 (2009).
4. Scott-Phillips, T. C. *Evolution* **68**, 1231–1243 (2014).
5. Krkosek, M., Lewis, M. A. & Volpe, J. P. *Proc. Biol. Sci.* **272**, 689–696 (2005).

CAREER DECISIONS

Of academic interest

A survey of 1,500 recent US biomedical PhD graduates found that white and Asian men, who are well represented in academic institutions, are most likely to show interest in becoming academic researchers (K. D. Gibbs *et al. PLoS ONE* **9**, e114736; 2014). Although scientists of all ethnicities reported losing interest in faculty careers as their doctoral studies continued, women's loss of interest was more pronounced, particularly for underrepresented minorities. The trend persisted after controlling for factors such as publication record and a sense of belonging. Understanding why career interests differ is crucial for increasing diversity, says study co-author Kimberly Griffin of the University of Maryland, College Park.

GRANTS

Paint me a picture

Applicants for US National Institutes of Health (NIH) grants will now be asked to contextualize past work, which could help early-career researchers. A revised ‘biosketch’ section instructs scientists to explain how their experience will allow them to accomplish the proposal's goals. Sally Rockey, the NIH's deputy director for extramural research, has said that the change puts focus on an applicant's skills and accomplishments, not just their publications. The format is being rolled out in January and will become mandatory for most NIH grants on 25 May. Online tools such as SciENcv have been updated to help applicants to create biosketches that can be used across various government agencies.

HIGHER EDUCATION

UK science numbers up

Between 2003 and 2012, UK student enrolment in the biological and physical sciences rose by more than 30% and overall university enrolment rose by more than 6%, according to a report by the country's Higher Education Statistics Agency. But across all fields, the number of postgraduate researchers increased by only 0.1% for 2012–13, the smallest increase since 2004. In the decade since 2003, enrolment at UK institutions fell sharply for students from the Middle East (–124%), Asia (–63%), Africa (–36%) and North America (–31%), but rose for students from Australasia (22%) and South America (21%).

THE DESCENT OF MAN

A precarious position.

BY CHRISTOPH WEBER

For one moment, the afternoon air suspends me 50 feet above the earth. For one moment, I am free.

An apple branch absorbs the shock of my landing. I dip my brush into the basket at my hip and pull out the black bristles, now dusted in yellow pollen. They say that bees were black and yellow, so I guess the colours are fitting.

I dab a white flower, dust some yellow on the stigmas, and move on — blossom to blossom, branch to branch.

The fruit trees grow tall here in Washington. The skinnies used to stunt them for easy picking, but after waking us, they tweaked the trees to utilize vertical space.

I leap. My muscled legs explode like pistons in the skinnies' useless old engines. I land on a stout branch; it flexes, then lobs me back into space. Free again.

I extend a thick, calloused hand, grasp the smooth red-barked limb in a handshake of trust, and redirect my swing. I do a little half turn, hope the other old ones are watching, and land backwards on the next branch.

"March!"

Someone was watching. Not an old one, though. It's Maestro, the skinny who names us all after the month he buys us.

"March, where is your rope?" he demands, staggering towards my tree, rifle in hand. Drunk, no doubt.

Tobacco juice hangs from his mouth like a black string. My lips curl in revulsion. He notices. He may be revolting, with a forehead big as the Moon, but he doesn't miss much.

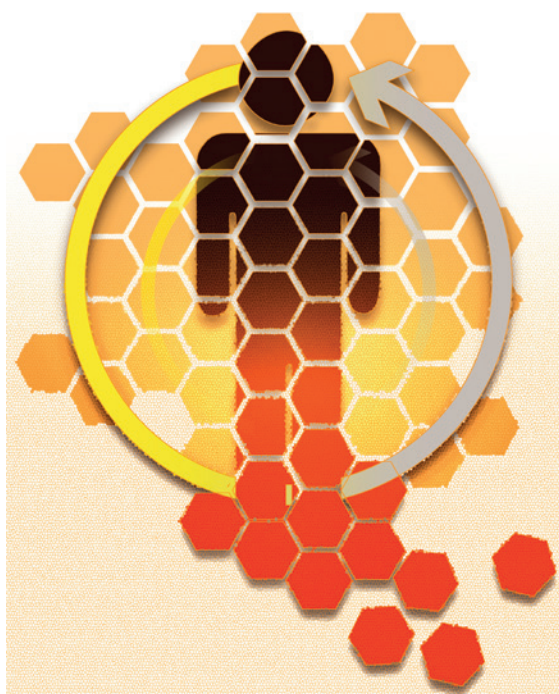
"Do I disgust you?" he asks, spitting a thick black rope.

I say nothing. Just size up his spindly arms. You may be tall, but I could break you in half, skinny.

The skinnies didn't wake us to climb trees. They woke us to cure their diseases, to fix their weak bones. They woke us for the same reason skinnies do everything — to help themselves. They'll tell you that it was out of some innate curiosity, but satisfying that curiosity is just another way they help themselves.

They were wrong about us — we're better climbers than they imagined, more

suitied to trees than they thought. We have the strength to climb and leap, and the bones to support rough landings, occasional falls. "More like the rest of Hominidae than we expected," they said. "Less like us than we thought." Like that's a bad thing. The skin-



nies used to enslave their own kind. When they needed labour to do the work of bees, it shouldn't have surprised anyone they'd turn to us.

There are some skinnies who oppose it, mostly in the South where it's too dry to grow much anymore. It's easy for them to talk about emancipation when they don't need us, not like the Northwest does. This is the last good growing land, and good growing requires pollinators. So here I am, 50 feet up in a tree with an ugly skinny yelling at me.

"March, I'm only gonna ask you one more time. Where is your rope!"

"Don't need it." I could climb this tree with my eyes closed.

Maestro hoots. "Don't need it! March, I own you — remember that! You fall from that tree, you damage *my* property. You know what I do to stonelickers that break my things?"

I stare back in silence.

"March, are you mean-muggin' me? Get down here. Now!"

I stand still. Maestro lifts his rifle. I know

I should come down, but in this small act of defiance I can see Maestro realizing that he doesn't *really* own me, and I can't help but grin. Anyway, he won't actually shoot me. I'm too useful.

The applewood next to my face explodes in a starburst of splinters. I wipe my eyes. Blink. Look down the barrel of Maestro's rifle. Dive. The wood reverberates with intercepted shot. I duck behind the trunk. Is Maestro aiming to kill? It doesn't matter. He's drunk and angry and the tree is exploding around me. His shots are close, too close. I stand with my back to the trunk and see April under her tree, looking up at me with wide, worried eyes.

I peek around the bole. Maestro's reloading. His head is down.

Something in the oldest, deepest recess of my mind comes alive and takes control. When my feet leave the powdery red bark, it's as if they belong to someone else. I slip into space and begin the descent of a hunter ambushing prey.

Maestro looks up just in time to see my heavy knees meet his face.

There is a crunching sound. A blossom of white pain in my legs. The impact was hard, even for my thick bones. I roll over. See Maestro's face. Understand the sound was not from my legs. I stagger to my feet as April and September Two run

over. September Two looks down at Maestro, shudders and picks up his rifle. April throws her arms around me, buries her face in my neck. When she looks up, her eyes glisten from under a strong brow, out from a face full of strength. A face full of pride. The face of an old one.

Old one — that's what we call ourselves. Skinnies like Maestro call us stonelickers. But if you find a decent skinny, which doesn't happen much, he might call us by another name.

Neanderthal.

The other old ones gather round and next thing I know we're running, all together, for the edge of the orchard where the dark forest beckons, offering concealment from catchers. We're running from skinnies, running into the ancient wood's embrace.

For the first time in our lives, we are running free. ■

Christoph Weber is an arborist in Reno, Nevada. Despite spending much of his life in trees, he is fully human.

ILLUSTRATION BY JACEY

➔ NATURE.COM

Follow Futures:

@NatureFutures

go.nature.com/mtoodm

47. Honda, Y. & Ishizuka, N. Organization of connectivity of the rat presubiculum: I. Efferent projections to the medial entorhinal cortex. *J. Comp. Neurol.* **473**, 463–484 (2004).
48. McNaughton, B. L., Battaglia, F. P., Jensen, O., Moser, E. I. & Moser, M.-B. Path integration and the neural basis of the 'cognitive map'. *Nature Rev. Neurosci.* **7**, 663–678 (2006).
49. Zhang, K. Representation of spatial orientation by the intrinsic dynamics of the head-direction cell ensemble: a theory. *J. Neurosci.* **16**, 2112–2126 (1996).
50. Redish, A. D., Elga, A. N. & Touretzky, D. S. A coupled attractor model of the rodent head direction system. *Network Comput. Neural Sys.* **7**, 671–685 (1996).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank A. Treves, J.-M. Fellous, M. Okun, A. Wallach, M. Geva-Sagiv and M. M. Yartsev for comments on the manuscript; B. Pasmantirer and G. Ankaoua for mechanical designs; S. Kaufman for bat training, assistance in neural recordings, and illustrations; T. Eliav and T. Tamir for help in experiments; A. Tuval and M. Weinberg for veterinary oversight; G. Brodsky for graphics; R. Eilam and C. Ra'anani for histology; and M. P. Witter for advice on reconstruction of tetrode-track locations

and anatomical delineations. This study was supported by research grants to N.U. from the European Research Council (ERC-NEUROBAT), the Human Frontiers Science Program (HFSP RGP0062/2009-C), the Israel Science Foundation (ISF 1017/08 and ISF 1319/13), and the Minerva Foundation; by a Clore predoctoral excellence fellowship to A.F.; and by an MIT-Israel (MISTI) student exchange internship to J.N.F. D.D. is the incumbent of the David and Inez Myers Career Advancement Chair in Life Sciences.

Author Contributions A.F., D.D. and N.U. designed the experiments and the analyses. A.F. performed the experiments, with contributions by D.D. and L.L. to some of the surgeries and tetrode recordings. A.F. and A.R. developed the toroidal model. A.F. and J.N.F. developed algorithms. A.F. analysed the data, and discussed with D.D., A.R., L.L. and N.U. the results and interpretations. A.F. and N.U. wrote the manuscript with input from D.D., A.R., J.N.F. and L.L.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to N.U. (nachum.ulanovsky@weizmann.ac.il).

METHODS

Subjects and behavioural training. Neural recording experiments were conducted in 3 different behavioural setups: (i) upright-crawling arena ($n = 4$ bats), (ii) vertical-ring arena ($n = 2$ bats), and (iii) flight room ($n = 1$ bat). All experiments were conducted in adult male Egyptian fruit bats, *Rousettus aegyptiacus* (weight 139–176 g at implantation). Bats were maintained on a reversed light/dark cycle; experiments were carried out in the dark phase. All experimental procedures were approved by the Institutional Animal Care and Use Committee of the Weizmann Institute of Science.

For the set of experiments conducted in crawling bats in setup number 1 (Figs 1–3), neural activity was recorded from the dorsal presubiculum of 4 bats, as they crawled in a square (50×50 cm) wooden box that was used throughout the training and recording period (Extended Data Fig. 1a). The square arena had 50-cm high walls, its floor and walls were coloured in black, and a portion of one of the walls was painted white and served as a polarizing cue (18-cm wide \times 50-cm high white cue card). The arena was placed horizontally on the floor of a large room; the position of the arena within the room was fixed throughout the duration of training and testing, and the locations of its four corners were calibrated periodically. Bats were trained to crawl on the arena floor that was covered with a soft rubber surface, in search for randomly scattered food (pieces of fruit: bananas or dates). During this random foraging task bats moved their head extensively in 3D, in azimuth, pitch, and roll (see Fig. 1d). All training and recording sessions were done under low-light conditions (illuminance 1 lx; measured using ILT-1700 illuminance meter, International Light, MA), a light level that allows these highly visual bats^{51,52} to clearly use vision.

The sequence of behavioural recording sessions was typically flanked by sleep sessions, during which the bat was comfortably placed inside a holding bag, with its head and neck outside the bag, and rested on a pedestal inside the experimental room. The full sequence of sessions was as follows: sleep, upright session 1, inverted session, upright passive session, upright session 2, sleep. Sleep sessions lasted ~ 10 min each, while the behavioural sessions lasted 15–25 min each. The total duration of the crawling experiment was ~ 2 h. In the ‘upright sessions’ the bat was required to crawl in the arena in search for randomly delivered food. In the ‘inverted session’ the bat was loosely held by the experimenter in a holding-bag and was moved above the arena with its head upside-down; while in the bag, the animal was free to move its head all around. In the ‘passive session’ the bat was placed in an upright position in the holding-bag, and was moved by the experimenter over the floor of the arena. Throughout the duration of the daily recordings, the bats were not taken outside of the recording room.

For the experiments in setup number 2, with bats crawling on a ring, we built a 90×50 cm box containing a vertically oriented ring, positioned 25 cm from the short wall of the box (Fig. 4a), and trained 2 bats to either crawl on the floor as before (Fig. 4a, pose 1), or climb onto the inner surface of the ring and complete a full 360° revolution (360° pitch angles; Fig. 4a, pose 2). The ring (50-cm diameter \times 8-cm width) was made from plastic mesh with 0.8×0.8 cm holes, which allowed the bat to cling to it and crawl along its inner surface. Importantly, we trained the bats to crawl in the same session both on the arena floor and on the vertical ring, so that the ring traversal was done of the bat’s own volition, and was intermingled with crawling on the arena floor. This design allowed comparing the neural activity during upright crawling on the arena floor, which had a limited range of pitch coverage (approximately $\pm 45^\circ$), with the activity on the ring, where behavioural sampling covered all possible pitch angles (360° pitch). In total, each session lasted for 30–45 min, with the bat spending roughly half of the time on the arena floor and half of the time on the ring. The recording session was flanked by two sleep sessions, lasting 10 min each.

For the flight experiments in setup number 3, one bat was trained to fly in a large flight-room ($5.8 \times 4.6 \times 2.7$ m; Fig. 5a) in search for randomly positioned food rewards (banana pieces). The flight room contained multiple objects, on which the reward was randomly placed. These objects included elevated poles and hanging balls, and the variety of objects and positions allowed for a maximal diversification of the flight trajectories, in order to produce sufficient coverage of the z axis and of all possible 3D directions. The recording session in the flight experiments typically lasted 45–60 min, and was flanked by two sleep sessions, lasting 10 min each.

No statistical method was used to predetermine sample size.

Surgery and recording techniques. All surgical and recording procedures were as described previously¹⁶. Briefly, after completion of training, bats were implanted with a four-tetrode lightweight microdrive (4-drive, Neuralynx; weight 2.1 g). Tetrodes (~ 45 μ m diameter) were constructed from four strands of platinum-iridium wire (17.8 μ m diameter, HML-insulated), bound together by twisting and melting their insulations; tetrode wires were gold-plated on the day before surgery to reduce their impedance to 300–700 k Ω . While the bat was under isoflurane anaesthesia, a circular opening (craniotomy of 1.8-mm diameter) was made in the skull over the right hemisphere. The centre of craniotomy was positioned over the dorsal presubiculum,

3.3 mm lateral to the midline and 3.15 mm anterior to the transverse sinus that runs between the posterior part of the cortex and the cerebellum (N. Ulanovsky, M. P. Witter and R. Eilam, *Stereotaxic Brain Atlas of the Egyptian Fruit Bat*, in preparation).

Following surgery, tetrodes were slowly lowered towards the presubiculum. The precise positioning of each tetrode was later reconstructed histologically (see below). The tetrodes were always moved at the end of each recording day (40–200 μ m daily), in order to obtain recordings from new ensembles of neurons daily. For each bat, one tetrode was left in an electrically quiet zone and served as a reference, and the remaining three tetrodes served as recording probes. During recordings in crawling bats, a unity-gain preamplifier (HS-18, Neuralynx) was attached to a connector on the microdrive. Signals were amplified ($\times 1,400$ – $5,000$) and band-pass filtered (600–6,000 Hz, Lynx-8 or DigitalLynx, Neuralynx), and a voltage-threshold was used for collecting 1-ms spike waveforms, sampled at 32 kHz. For flight experiments, we used a wireless neural-recording device (an improved version of the device described in ref. 36). In all setups, data were collected continuously throughout all the behavioural and sleep sessions of each recording day (1.5–2 h daily).

Video recordings and reconstruction of 3D head-direction angles. In order to reconstruct the 3D head direction of the bat as it was crawling in setup number 1 (Extended Data Fig. 1a), using the 2D tracking data of a single camera, we added a custom 3D tetrahedral arrangement of four LEDs to the recording head-stage on the bat’s head (illustrated in Extended Data Figs 1b, c and 11). The positions of these LEDs were recorded using a single camera that was placed vertically overhead the arena, and was connected to a video-tracker (Neuralynx) with a 25-Hz sampling rate, which was synchronized to the neural recordings. Combined with the data-analysis algorithm described below, this LED arrangement enabled a precise mapping from the position of these four LEDs in the camera image, onto a complete set of coordinates describing the five degrees of freedom of the bat’s head. These degrees of freedom include the two spatial coordinates (x, y) for the position inside the arena, and three Euler angles describing the head rotation: yaw (‘azimuth’, ϕ), pitch (θ), and roll (γ). For more details, see Extended Data Figs 11 and 12 and the last section of the Methods, which describes the algorithm for reconstructing the 3D head direction.

In the ring apparatus (setup number 2; Fig. 4a), when the bat crawled on the floor in the upright position (Fig. 4a, pose 1), the 3D head direction was measured using the video data acquired from the overhead camera (Fig. 4a, ‘cam 1’), as described above for setup number 1. When the bat crawled on the inside part of the ring (Fig. 4a, pose 2), we used instead a second camera pointing horizontally (Fig. 4a, ‘cam2’) to measure the pitch (θ) of the bat’s head over the entire 360° of possible pitch angles, based on the red–green LED pair. In addition, we used the horizontally pointing camera to determine whether the bat performed a ring traverse in the east-bound or westbound direction (see Extended Data Fig. 6).

In the flight apparatus (setup number 3; Fig. 5a), we estimated the head direction of the bat by taking advantage of a previous finding that in flying Egyptian fruit bats, heading-direction (flight direction) is tightly coupled to head direction during flight⁵³. We computed the head-direction azimuth (ϕ) and pitch (θ) of freely flying bats, based on the heading-direction, using the following equations:

$$\phi = \text{angle}(\Delta x + \Delta y \cdot i)$$

$$\theta = \text{angle}(\sqrt{\Delta x^2 + \Delta y^2} + \Delta z \cdot i)$$

Where x, y, z are the position of the bat’s head in 3D space; $\Delta x, \Delta y, \Delta z$ are the changes in head-position between consecutive video frames; and i is the imaginary number. The 3D head-position (x, y, z) was recorded at 25-Hz using two synchronized video-cameras placed at two of the upper corners of the flight-room (Fig. 5a), which tracked the position of LEDs connected to the wireless head-stage on the bat’s head, enabling 3D reconstruction of the animal’s 3D head position (see ref. 36). The 3D head position was mildly smoothed using a cubic smoothing spline; $\Delta x, \Delta y, \Delta z$ were computed from these smoothed values.

Spike sorting and criteria for cell inclusion. All spike-sorting procedures were identical to those described previously^{9,16,54}. Briefly, spike waveforms were sorted on the basis of their relative energies and amplitudes on different channels of each tetrode (software: SpikeSort3D, Neuralynx; see example in Extended Data Fig. 1). Data from all sessions, including the two sleep sessions, were spike-sorted together. Well-isolated clusters of spikes were manually encircled (‘cluster-cutting’), and a refractory period (< 2 ms) in the interspike-interval histogram was verified. In the upright-crawling experiments (setup number 1, Figs 1–3), a total of 266 well-isolated neurons were recorded from dorsal presubiculum of 4 bats. Of these 266 neurons, we included for analysis a total of 122 cells, based on the following criteria: (i) the cell fired at least 30 spikes in each of the upright behavioural sessions; (ii) the cell

was directionally stable between the two upright sessions, with the Pearson correlation coefficient between the one-dimensional (1D) neural tuning curves for azimuth, pitch, or roll being higher than 0.5 ($r \geq 0.5$).

In the ring-experiments (setup number 2, Fig. 4), a total of 180 well-isolated neurons were recorded from the dorsal presubiculum of 2 bats. Of these neurons, we included for analysis 175 cells, based on the following criteria. (i) For computing the tuning curves for pitch, we included only cells that fired at least 25 spikes on the ring during ring-crawling epochs. (ii) For computing the tuning curves for azimuth, we included only cells that fired at least 25 spikes on the floor during floor-crawling epochs. (iii) For comparing pitch tuning on the ring with that on the floor (Fig. 4b, d), and for computing the 2D tuning curves for pitch \times azimuth (Fig. 4h–k), we included only cells that fired at least 25 spikes on the ring and at least 25 spikes during floor-crawling.

In the flight experiments (setup number 3, Fig. 5), a total of 72 well-isolated neurons were recorded from the dorsal presubiculum of one bat. We included for analysis a total of 67 cells, which fired at least 30 spikes during the flight epochs.

Head-direction tuning curves. The head-direction tuning of a neuron was computed as follows: for azimuth tuning, we binned the data in 18 bins of 20° each, and computed the firing rate in each bin by dividing the number of spikes in that bin by the time the animal spent in that bin; to compute the neuronal tuning to pitch and roll in the upright-crawling experiments (Figs 1–3), we binned the data in 10 bins of 10° each, with each bin labelled according to its centre angle; the extreme bins (labelled as -45° and $+45^\circ$) contained angles smaller than -40° or larger than $+40^\circ$, respectively. The resulting 1D tuning curve was smoothed using a rectangular window of size 3 (the smoothing was done on the firing-rate curve, that is, after dividing the unsmoothed spike-curve by the unsmoothed time-spent curve). For azimuth data, the smoothing was applied in a circular manner. Unvisited bins (where total time-spent, before smoothing, was < 2 s) were discarded from analysis. The peak firing rate was defined as the highest rate in the head-direction tuning curve. Tuning width was defined as the width of the tuning curve at 25% of the maximum height ('full tuning width'), where the height was measured from the peak firing rate to the minimal firing rate.

The directionality of the tuning curve in azimuth in the upright-crawling experiments was quantified by computing the Rayleigh vector length of the circular distribution, using the following definition⁵⁵:

$$\text{Rayleigh vector length} = \frac{\pi}{n \cdot \sin(\frac{\pi}{n})} \sum_{j=1}^n r_{\varphi_j} e^{-i\varphi_j} / \sum_{j=1}^n r_{\varphi_j}$$

Where n is the number of circular head-direction bins, φ_j is the direction in radians of the j -th circular bin (namely, $2\pi j/n$), and r_{φ_j} is the average firing rate given the animal's head direction.

In the upright-crawling experiment (setup number 1) we applied a shuffling-based significance test^{15,16,25,26}, in order to identify neurons that were significantly tuned to azimuth, pitch, or roll. To determine significance in azimuth, we performed the shuffling analysis on the Rayleigh vector length. As pitch and roll were limited in their range and therefore non-circular, we could not use the Rayleigh vector length; instead, to determine significance for pitch and roll, we applied shuffling analysis for the tuning width. Significance was determined for each neuron separately; specifically, for each recorded neuron, the entire sequence of spikes was time shifted by a random (uniformly distributed) time interval, with the end of the session wrapped to the beginning. This preserved the spike number and the temporal structure of the neuron's firing pattern, but dissociated the time of spiking from the animal's actual behaviour. This procedure was repeated 2,000 times for each neuron. A session was defined as tuned to azimuth if its Rayleigh vector length exceeded the 95th percentile of the shuffled distribution for this session. Similarly, a session was defined as tuned to pitch or roll, if the actual tuning width for pitch or roll was narrower than the 95th percentile of the shuffled distribution for this session.

To determine whether a neuron was tuned conjunctively to more than one Euler angle, we first computed the tuning significance of each session for each of the Euler angles independently, based on the shuffling test for the relevant head-direction angle; then, we defined significant 'conjunctive' or 'pure 1D' tuning by taking into account the two upright behavioural sessions, and computing their average 'surprise' function:

$$\text{Average surprise} = -[\log(P_{\text{session 1}}) + \log(P_{\text{session 2}})]/2$$

where $P_{\text{session 1}}$ and $P_{\text{session 2}}$ are the P values obtained from the shuffling-based significance tests for upright sessions 1 and 2. A neuron was categorized as directional in azimuth, if its average surprise function in azimuth was ≥ 1.301 (corresponding to significance at $\alpha = 0.05$); similar definition was used to categorize neurons as directional in pitch or roll. If the neuron was significantly directional in

more than one head-direction angle, for example, in both azimuth and pitch, it was categorized as a 'conjunctive azimuth \times pitch' cell (for example, Fig. 2d, e). If the neuron was directional only in one head-direction angle, for example, only in azimuth, it was categorized as a 'pure azimuth' cell (for example, Fig. 2b).

Directional tuning in the inverted session was assessed, as in the upright case, using shuffling analysis of the Rayleigh vector length (see Extended Data Fig. 5a). We also analysed the non-significant cells by dividing two non-overlapping groups, those that were 'weakly tuned' under inversion (that is, non-significant, but still clearly tuned) versus truly 'untuned cells' (see Extended Fig. 5d, e). These two groups were separated according to the goodness of fit (R^2) to circular normal function (von Mises function, see below), higher versus lower than $R^2 = 0.45$.

In the ring experiment (setup number 2), azimuth-significant and pitch-significant cells were defined based on shuffling analysis of the respective Rayleigh vector length. Azimuth cells were defined as neurons with significant Rayleigh vector length based on the tuning curve they had on the arena floor (Fig. 4a, pose 1). Pitch cells were defined as neurons with significant Rayleigh vector based on their tuning curve on the ring (Fig. 4a, pose 2). In both cases, Rayleigh vector length and shuffling analysis were similar to those conducted for azimuth cells in the upright-crawling experiment number 1 (see above). In addition, in order to increase the sample size in the ring experiment, we included cells that did not pass the shuffling threshold but had Rayleigh vector length > 0.5 . The tuning width of pitch cells in the vertical ring experiment was defined, similarly to azimuth cells, as the width of the tuning curve at 25% of the maximum height, after subtracting the baseline (the minimal firing rate). The analysis of azimuth tuning on the ring (Extended Data Fig. 6) was restricted to azimuth cells whose preferred direction on the arena floor (Extended Data Fig. 6a, b, pose 1) was aligned to the cardinal axis of the ring (east–west, deviating from this axis $< \text{half of the mean tuning width}$). The movement of the inverted bat could be either in the preferred azimuth of the cell (Extended Data Fig. 6a), or in the exact opposite (180°) direction (Extended Data Fig. 6b).

Significant azimuth and pitch cells in flight (setup number 3) were defined in a similar way to their definition in the upright-crawling experiment number 1, based on shuffling analysis of the Rayleigh vector length for azimuth cells, and based on shuffling analysis of the tuning width for pitch cells.

In Fig. 4e–h, we fitted the 1D head-direction tuning curves of neurons with a 1D circular normal function, known also as von Mises function, which has the following form:

$$R_i(\varphi) = c_1 e^{\kappa \cos(\varphi - \varphi_i)} + c_2$$

Where φ is the preferred direction of this unit in radians, φ_i is the direction in radians of the i -th circular bin, and κ , c_1 , and c_2 are constants. The quality of the von Mises curve-fitting to the experimental data was used also to assess the unimodality of the neuronal 1D tuning curves (Fig. 4f).

In Fig. 4i, we fitted the 1D azimuth (φ) and 1D pitch (θ) tuning curves of conjunctive neurons with 1D von Mises functions, and combined these 1D functions to generate a 2D von Mises function (Fig. 4i, left):

$$R_i(\varphi, \theta) = c_1 e^{[\kappa_1 \cos(\varphi - \varphi_i) + \kappa_2 \cos(\theta - \theta_i)]} + c_2$$

2D head-direction firing-rate maps. To analyse the conjunctive directional firing-rate maps of presubiculum neurons, we extended the 1D binning used in the section on 'head-direction tuning curves', to 2D head-direction maps (shown in Fig. 2a–e; Extended Data Fig. 3b–e, top; and Extended Data Figs 7 and 8). To construct these 2D maps, we used 18 bins for azimuth, 10 bins for pitch and 10 bins for roll (thus, an azimuth \times pitch map included 18×10 bins, while a pitch \times roll map included 10×10 bins). For each combination of rotation angles, we computed two maps: (i) the time-spent in each 2D directional bin, and (ii) the spike-count per bin. These two maps (time-spent and spike-count) were then individually smoothed using a Gaussian kernel with standard deviation of $\sigma = 1.7$ bins; circular (wrapped) smoothing was applied only for the azimuth in the azimuth \times pitch maps and azimuth \times roll maps. The 2D head direction firing-rate map was then computed for each neuron by dividing bin-by-bin the two smoothed maps of spike count and time spent. Unvisited bins (where total time-spent, before smoothing, was < 0.5 s) were discarded from analysis, unless they were adjacent to a visited bin (with a total time-spent ≥ 0.5 s), and are shown as white bins on the 2D plots. Exclusion based on minimal time spent was different in 1D versus 2D firing maps (2 s versus 0.5 s), because the average time spent in each bin was lower for 2D maps, due to the increase in the dimensionality of the data. The peak firing rate was defined as the highest observed firing rate in any of the bins of the 2D firing-rate map. The 2D maps were plotted with colours normalized between the minimal and maximal firing rate. It is important to note that these 2D maps were plotted for visualization only, while all the statistical tests used for determining the cell category (for example, pure-azimuth cell,

conjunctive azimuth \times pitch cell, etc.) were done strictly on the 1D tuning-curves, as described above.

We also plotted 3D firing-rate maps for all 3 Euler angles conjunctively, in a Cartesian representation (azimuth \times pitch \times roll), as shown in Fig. 2f. These 3D maps were computed in the same way as described in ref. 36 for the case of 3D place cells.

Spatial firing-rate maps and their quantification. To analyse the 2D positional firing-rate maps of presubiculum neurons (Extended Data Fig. 3b–e, bottom), we divided the recording arena into 3.5×3.5 cm square bins, and computed a time-spent map and a spike-count maps for each neuron. These two maps were individually smoothed using a Gaussian kernel with standard deviation of $\sigma = 1.7$ bins. The 2D positional firing-rate map was then computed for each neuron by dividing bin-by-bin the two smoothed maps of spike-count and time-spent. Unvisited bins (where total time-spent, before smoothing, was < 0.5 s) were discarded from the analysis, unless they were adjacent to a visited bin (with a total time-spent ≥ 0.5 s), and are shown as white bins on the 2D plots.

The spatial information, in bits per spike, and the sparsity index, were computed from the smoothed firing-rate map of each cell, as described previously^{9,16,56,57}:

$$\text{Spatial information (bits/spike)} = \sum p_i(r_i/\bar{r}) \log_2(r_i/\bar{r})$$

$$\text{Sparsity} = \langle r_i \rangle^2 / \langle r_i^2 \rangle = (\sum p_i r_i)^2 / \sum p_i r_i^2$$

where r_i is the firing rate of the cell in the i -th bin of the place-field, p_i is the probability of the animal being in the i -th bin (time spent in i -th bin/total session time), \bar{r} is the overall mean firing rate, and i is running over all the visited bins. The spatial-information index is high for neurons with high spatial selectivity. The sparsity index is bound between 0 and 1, with high values indicating poor spatial selectivity.

Angular velocity tuning curves. The angular velocity tuning of a neuron was computed as the time derivative of each of the head-direction angles:

$$\text{Azimuth velocity} = \frac{d\varphi}{dt}$$

$$\text{Pitch velocity} = \frac{d\theta}{dt}$$

$$\text{Roll velocity} = \frac{d\gamma}{dt}$$

For azimuth, the computed angular velocity of the head was binned into 8 bins, each of 25° s^{-1} (total range between $\pm 100^\circ \text{ s}^{-1}$); for pitch and roll we binned the velocity data into 8 bins, each of 20° s^{-1} (total range between $\pm 80^\circ \text{ s}^{-1}$). We then computed the firing rate in each bin by dividing the number of spikes in that bin by the time the animal spent in that bin. The resulting curve was smoothed using a rectangular window of size 3. Unvisited bins (where total time spent, before smoothing, was < 0.5 s) were discarded from analysis.

Toroidal representation and correlations between the directional firing-rate maps in the upright versus inverted positions. The two axes of the torus correspond to:

φ (azimuth): the angle of the inter-aural (ear-to-ear) axis projection onto the x - y plane. Notably, this definition of the azimuth in toroidal coordinates results in it being insensitive to rotations in pitch, and thus naturally captures the azimuth direction in both upright and inverted positions.

θ (pitch): the angle of the rostro-caudal axis of the head in the plane which is defined by the z -axis and the projection of the rostro-caudal axis onto the x - y plane. This angle is the common definition of pitch, with the exception that in the toroidal coordinates it is continuous with a cyclical range of $\pm 180^\circ$ (instead of $\pm 90^\circ$).

In order to compare spherical vs toroidal representations, we used the following transformation of Euler angles from spherical to toroidal representation:

$$\text{If position} = \text{upright } \varphi \rightarrow \varphi, \theta \rightarrow \theta$$

$$\text{If position} = \text{inverted } \varphi \rightarrow \varphi + 180^\circ, \theta \rightarrow 180^\circ - \theta$$

And from toroidal to spherical representation:

$$\text{If } -90^\circ < \theta < 90^\circ \varphi \rightarrow \varphi, \theta \rightarrow \theta, \text{ position} = \text{upright}$$

$$\text{If } 90^\circ < \theta < 270^\circ \varphi \rightarrow \varphi + 180^\circ, \theta \rightarrow 180^\circ - \theta, \text{ position} = \text{inverted}$$

Specifically, in order to compare spherical versus toroidal representations, we first computed the 2D azimuth \times pitch directional rate-maps for the upright and inverted sessions in spherical coordinates, as described above. Then we transformed them into toroidal coordinates, where the outer surface of the torus represents the 2D map for the upright session, while the inner surface represents the 2D map for the inverted session after it underwent two transformations: a shift of all bins along the azimuth dimension by 180° ; and flipping the pitch axis. The rationale behind these two transformations is based on two key features underlying the toroidal model:

(i) The first feature is that the azimuth direction of the head in the inverted bat is an extension of the azimuth direction in the upright position (after a shift of 180°), and therefore the upright and inverted positions are not two distinct states but rather a continuum. Therefore, pitching the head at angles greater than $+90^\circ$ will result in flipping the bat from upright to inverted position, and further rotation along the pitch axis will eventually flip the bat back to the upright position, without changing the toroidal azimuth (in contrast, in the spherical model, such a change in pitch will change the azimuth by 180° ; compare Fig. 3d to 3c; azimuth values are in blue). For example, the response of a cell to the east direction when the animal is in upright position is given on the outer part of the torus, while the inner side represents the response of the cell to the west direction when the animal is inverted (Extended Data Fig. 7c and right panel of 7d). Therefore, the first step required to transform from spherical to toroidal representation, is shifting all the bins of the 2D directional map of the inverted session (which is computed initially in spherical coordinates) by 180° in azimuth, in order to re-align each azimuth-direction of the inverted session (inner part of the torus) with its equivalent azimuth in the upright session (outer part of the torus).

(ii) The second feature of the toroidal model is that the pitch angle of the head is circular and spans a range of 360° (Fig. 3d), and is computed in the laboratory reference frame (allocentric representation). In order to represent the pitch continuously in the laboratory reference frame, we concatenated the $[-90^\circ$ to $+90^\circ]$ pitch axis of the upright session with the $[+90^\circ$ to $-90^\circ]$ flipped pitch axis of the inverted session, to create a full 360° circular pitch axis. In this way, pitch bins that correspond to adjacent head-pitch in the upright and inverted sessions (adjacent in absolute space, laboratory reference frame) are plotted adjacently to each other on the toroidal manifold.

To test whether the neuronal data are consistent with the toroidal representation, we computed the Pearson correlation coefficient, r , between the 2D directional rate-maps (azimuth \times pitch) for each upright session versus the inverted session of the same neuron, under various transformations. We started by comparing the 2D map for the upright session (Extended Data Fig. 7e) with the map for the inverted session before any transformations (Extended Data Fig. 7f), where pitch was represented egocentrically (see Extended Data Fig. 9), relative to the body-reference frame, and the azimuth was not realigned by 180° shift (Extended Data Fig. 7f). Next we compared the 2D map for the upright session versus several angular transformations of the inverted-session map. These transformations of the inverted session included: (i) an azimuth shift of 180° (Extended Data Fig. 7g); (ii) flipping the pitch axis, in order to convert into an allocentric representation (Extended Data Fig. 7h); and (iii) both an azimuth shift of 180° and flipping the pitch (Extended Data Fig. 7i).

Correlation between the upright and the inverted session was computed for the spherical versus the toroidal representation for all cells that were significantly tuned to azimuth or pitch (Fig. 3e). In spherical representation, this correlation was computed between the upright session (for example, Extended Data Fig. 7e) and the inverted session after transformation (ii) above (for example, Extended Data Fig. 7h). In toroidal representation, this correlation was computed between the upright session (Extended Data Fig. 7e) and the inverted session after transformation (iii) above (for example, Extended Data Fig. 7i).

To assess which of these angular transformations is more consistent with the neural data, we computed the increase in correlation (Δcorr , see Extended Data Fig. 7k), defined as the difference in correlation coefficient r before and after applying the indicated angular transformations on the inverted map. For example, to assess how much an azimuth shift of 180° improved the correlation between the upright-session map and inverted-session map, we computed the difference between the correlation r indicated in Extended Data Fig. 7g and the correlation indicated in Extended Data Fig. 7f; and similarly for the other angular transformations. These correlation differences were computed separately for non-overlapping sets of neurons categorized as pure azimuth, pure pitch, or azimuth \times pitch cells (see Extended Data Fig. 7k). Analogous calculations were conducted for the head-roll (Extended Data Fig. 7l).

Histology, and functional-anatomical gradients of head-direction cells. Histology was done as described previously¹⁶. In brief, tetrodes were not moved after the final recording session. The bats were anaesthetized, and in some bats electrolytic lesions (DC positive current of $30 \mu\text{A}$, 15-s duration) were done to assist in the precise reconstruction of tetrode positions. Lesions were made every $500 \mu\text{m}$

starting from the most ventral recording position. Immediately after doing the lesions, the bat was given an overdose of sodium pentobarbital and, with tetrodes left *in situ*, was perfused transcardially using phosphate buffer saline followed by fixative (4% paraformaldehyde + 0.1 M phosphate buffer saline). The brains were removed and stored in fixative. Subsequently, the right hemisphere was embedded in paraffin, and 10- μ m sagittal sections were cut; every third section was mounted on glass slides, resulting in 30- μ m intervals between adjacent mounted slides. The sections were then Nissl-stained and coverslip added, and a light microscope fitted with a digital camera was used to determine tetrode placement in the dorsal presubiculum.

The anatomical localization of the recorded head-direction cells was determined by the positions of the tetrodes based on the Nissl sections, together with the documentation of their recording depths. The position along the anterior-posterior axis was measured by projecting the tetrode-track onto a curved line going through the middle of the dorsal-ventral extent of the dorsal presubiculum (an example is shown in Fig. 2h, yellow line). The medio-lateral position of the tetrode track was determined by comparing the recording location to a set of reference sagittal sections with known medio-lateral coordinates (courtesy of M. P. Witter).

For the 1D functional map plots (Extended Data Fig. 4b–e), the percentage of each cell category (pure azimuth, pure pitch, pure roll, or all conjunctive cells) on each tetrode track was calculated out of the total number of directionally significant cells on that tetrode-track. In Extended Data Fig. 4b–e, only tetrode-tracks with $n > 5$ significant cells per track were included in this analysis.

For the 2D functional map plots (Extended Data Figs 4f–h, j, k, m, n), we calculated the percentages based on all the $n = 78$ significant head-direction cells (and not per-tetrode-track as done in Extended Data Fig. 4b–e). Spatial bins of 0.2×0.2 -mm were used, and an adaptive-binning procedure was applied to increase the bin-size in areas where recordings were sparser. Bins that were further than 500 μ m from any significantly-tuned cell in Extended Data Fig. 4f were whitened in Extended Data Fig. 4g, h, j, k, m, n, and were excluded from the analysis. In Extended Data Fig. 4f we plotted all the 266 cells recorded in setup number 1, for display purposes; note that only the significant head-direction cells, indicated by coloured dots, went into computing the 2D maps in Extended Data Fig. 4g, h, j, k, m, n. In Fig. 2i, all the $n = 78$ significant head-direction cells were included, because that analysis was not done per tetrode-track (unlike Extended Data Fig. 4b–e).

Algorithm for reconstructing the 3D head direction (three Euler angles of the head) using one overhead camera. In order to reconstruct the 3D head direction of the bat from 2D tracking data of a single camera, we added a custom 3D tetrahedral arrangement of four LEDs to the recording head-stage on the bat's head (illustrated in Extended Data Figs 11 and 12a). The positions of these LEDs were recorded using a single camera that was placed vertically overhead the arena, and was connected to a video-tracker (Neuralynx) with sampling rate of 25 Hz, which was synchronized to the neural recordings. Combined with the data analysis algorithm described below, this LED arrangement enabled a precise mapping from the position of these four LEDs in the camera image, onto a complete set of coordinates describing the five degrees of freedom of the bat's head. These degrees of freedom include the two spatial coordinates (x, y) for the position inside the arena, and three Euler angles describing the head rotation: yaw ('azimuth', φ), pitch (θ), and roll (γ). Note that in principle there is a degeneracy through which it is impossible to tell whether a bat is upside down; in practice, the context of the experiment resolved this ambiguity, as in our main experiments (see main Figs 1–3) we conducted separate recording sessions for bats crawling upright or being inverted upside-down. To track the inverted position, we placed a mirror on the arena box and tracked the reflections of the LED-array in the mirror, using the overhead camera. Because the LEDs were positioned on a Microdrive on the posterior–dorsal part of the head, opposite from the bat's eyes, the bat could not see the reflections of the LEDs in the mirror. The usage of a mirror allowed us to employ the exact same reconstruction algorithm for 3D head direction as we used in the upright condition; the only change that we did to the algorithm was to adjust it carefully in order to correct for the mirror-symmetry flip of the LEDs' reflections (as seen by the camera) during the inverted session. This adjustment was verified in a calibration session in which we compared upright angles to angles of an inverted headstage that was tracked through the mirror. As in the upright tracking, the head-direction angles in the inverted position were calculated relative to the laboratory reference frame.

For the ring experiments (main Fig. 4) we measured the 3 Euler angles using the overhead camera, as described below, when the bat was crawling on the arena floor (Fig. 4a, pose 1); in contrast, when the bat locomoted on the ring itself (Fig. 4a, pose 2), we measured only the pitch angle of the head using the side camera, and did not measure the other Euler angles (see above for more details about video-tracking in the ring experiments).

Notably, the Euler angles need to be measured with respect to the arena. For example, φ is the angle between the projection of the main axis onto the (x, y) plane (that is, onto the floor of the arena) and between the x axis of the arena coordinate system, as shown in Extended Data Fig. 12a, and it corresponds to the azimuth

(yaw) of the head. Similarly, θ is the angle between the main axis and the (x, y) plane of the arena, as illustrated in Extended Data Fig. 12b; θ corresponds to the bat elevating its head upwards or downwards (pitch angle of the head). γ measures the rotation of the device around the main axis, and corresponds to a sideways roll (tilt of the head), as illustrated in Extended Data Fig. 12a. The algorithm extracts these Euler angles from the 2D tracking data in four steps:

1) Derivation of the Euler angles with respect to the plane of the camera ($\varphi_c, \theta_c, \gamma_c$) from the image of the 2D projections of the LEDs in the camera coordinate frame (x_c, y_c); see Extended Data Fig. 12a–c.

2) Reconstruction of the 3D position of the four LEDs in the camera coordinate frame (x_c, y_c, z_c), based on the Euler angles ($\varphi_c, \theta_c, \gamma_c$) calculated in step 1.

3) Transformation of the 3D positions of the LEDs from the camera coordinates into arena coordinates (x_A, y_A, z_A); see Extended Data Fig. 12d.

4) Derivation of the Euler angles with respect to the arena ($\varphi_A, \theta_A, \gamma_A$), computed from the projection of the positions of the 3D LEDs in the arena coordinates onto the arena x - y plane.

φ_A, θ_A , and γ_A are the Euler angles in the arena coordinates, which we computed here; these are the head-direction angles (φ , azimuth; θ , pitch; γ , roll) reported in this study. Note that these angles are given in the absolute room coordinates (allocentric reference frame). The detailed calculations of each step are described below.

To validate this algorithm, we mounted the LED array onto a surgical stereotax (Kopf), which we can rotate and set the Euler angles to known 'ground truth' values, and took video data at various locations (x, y) in the arena, while varying all the three Euler angles in discrete steps over the physiologically relevant range. The comparison of the actual experimentally set angles (as set in the stereotax) and the read-out from the algorithm showed a reliable reconstruction, with a mean absolute error of 5.6° .

Following are details of the 4 steps used by the algorithm to extract the 3 Euler angles:

Step 1: one aim of the 3D LED apparatus was to enable estimating the Euler angles in the camera frame based purely on the ratios of measured quantities, as this renders the procedure robust against a change of scale (for example, when the animal moves its head up and down, that is, closer or farther from the camera).

By arranging the four LEDs in two offset perpendicular lines (a tetrahedral arrangement, as depicted in Extended Data Fig. 11 and 12a), the relative intersection point between the two lines in the image becomes a function of θ_c and γ_c , while φ_c can be determined using the projection of the main axis ($\mathbf{P}_1, \mathbf{P}_2$) onto the plane of the arena as seen from the camera ('the camera plane'). Based on this, the camera angles can be derived using the following relations: we define the geometry of the LED arrangement as shown in Extended Data Fig. 12a, where f_1 is the distance from the centre of the LED arrangement along the main axis to the front red LED, while f_2 is the distance from the centre to the rear green LED; h is the height between the main top axis and the transverse axis; r_1 and r_2 are the distances from the centre to the right and to the left blue LEDs, respectively; a_1, a_2, b_1 , and b_2 are the distances of the projection along the camera plane of the 4 arms of the LED arrangement (that is, front, rear, right, and left, respectively), as defined in Extended Data Fig. 12c (the dashed blue and dashed orange lines in Extended Data Fig. 12c denote the projections onto the camera plane of the LEDs and the segments connecting them, and are plotted also in Extended Data Fig. 12a; the dashed orange line is plotted also in Extended Data Fig. 12b).

α_c is defined as the distance of one of the blue LEDs from the intersection point, in the camera plane, normalized to the distance between the two blue LEDs in the camera plane:

$$\alpha_c = b_1 / (b_1 + b_2) = [r_1 \cos(\gamma_c) - h \sin(\gamma_c)] / [(r_1 + r_2) \cos(\gamma_c)] \quad (1)$$

We now define the subscript '0' as the corresponding ratio in the neutral position ($\varphi_c = \theta_c = \gamma_c = 0$), which can be determined from the known geometry of the device. Thus α_0 is the same normalized distance in the limiting case of $\gamma_c = 0$:

$$\alpha_0 = r_1 / (r_1 + r_2) \quad (2)$$

From (1) and (2) it follows that:

$$\alpha_0 - h \tan(\gamma_c) / (r_1 + r_2) = \alpha_c \quad (3)$$

From (3) we can derive γ_c :

$$\gamma_c = \arctan [(\alpha_0 - \alpha_c) (r_1 + r_2) / h] \quad (4)$$

Similarly, β_c is defined as the distance of the red LED from the intersection point in the camera plane, normalized to the distance between the red and green LEDs in the camera plane:

$$\beta_c = a_1 / (a_1 + a_2) = [f_1 \cos(\theta_c) - h' \sin(\theta_c)] / [(f_1 + f_2) \cos(\theta_c)] \quad (5)$$

where:

$$h' = h / \cos(\gamma_c) \quad (6)$$

given:

$$\beta_0 = f_1 / (f_1 + f_2); \quad (7)$$

Where, as before, the subscript '0' defines the corresponding ratio in the neutral position. See Extended Data Fig. 12b for illustration of the angle θ_c .

From (5) and (7) it follows that:

$$\beta_0 - h' \tan(\theta_c) / (f_1 + f_2) = \beta_c \quad (8)$$

Finally, from (6) and (8) we derive θ_c :

$$\theta_c = \text{atan} [(\beta_0 - \beta_c) (f_1 + f_2) \cos(\gamma_c) / h] \quad (9)$$

To derive φ_c :

$$\mathbf{v} = \mathbf{P}_1 - \mathbf{P}_2 \quad (10)$$

Where \mathbf{P}_1 and \mathbf{P}_2 are the projections of the red and green LEDs on the camera plane, respectively (see Extended Data Fig. 12a).

It follows that:

$$\varphi_c = \text{atan}[v_y / v_x] + \pi H(-v_x) \quad (11)$$

Where v_x and v_y are the x and y components of the vector \mathbf{v} , $H()$ is the Heaviside step-function, and $\text{atan}[x]$ is assumed to be the principal branch of tan, with $\text{atan}[x]$ between $-\pi/2$ and $\pi/2$.

Taken together, the Euler angles derived here (φ_c , θ_c , γ_c) are computed in the camera frame, based on the position of the LEDs in the camera plane.

Step 2: having obtained the Euler angles in the camera frame in step 1, the 3D position of the four LEDs can now be calculated in camera coordinates by applying a rotation matrix with the obtained Euler angles, in order to rotate the device from the neutral 'zero' orientation (that is, non-tilted, upright position) to the view in the camera frame:

$$(x_c, y_c, z_c) = \text{ROT}(\varphi_c, \theta_c, \gamma_c) (x_0, y_0, z_0) \quad (12)$$

where (x_0, y_0, z_0) is the configuration of the four LEDs in the neutral orientation, (x_c, y_c, z_c) is the configuration of the four LEDs in the camera frame, and $\text{ROT}(\varphi_c, \theta_c, \gamma_c)$ is the standard rotation matrix defined by the Euler angles derived in step 1.

Step 3: when the bat is in the centre of the arena, straight under the camera (position (0,0)), the camera and arena frames are perfectly aligned. However, in general the device is off-centre, and the difference between the two frames needs to be accounted for. An example of this mis-alignment of the two coordinate frames is shown in Extended Data Fig. 12d. In order to obtain the position of the LEDs in arena coordinates (x_A, y_A, z_A) , the basis vectors of the camera frame \mathbf{X}_C , \mathbf{Y}_C and \mathbf{Z}_C

need to be expressed in arena coordinates, that is: \mathbf{X}_A , \mathbf{Y}_A and \mathbf{Z}_A (Extended Data Fig. 12d). By construction, \mathbf{Z}_C is in line with the line-of-sight from the centre of the device to the camera. The camera plane is perpendicular to \mathbf{Z}_C and is spanned by \mathbf{X}_C and \mathbf{Y}_C , where \mathbf{X}_C is the normalized projection of \mathbf{X}_A onto the camera-plane. From requiring \mathbf{Y}_C to be perpendicular to both \mathbf{X}_C and \mathbf{Z}_C , it follows that:

$$\mathbf{Y}_C = \mathbf{X}_C \times \mathbf{Z}_C \quad (13)$$

Having obtained the local coordinate transformation connecting arena and camera coordinates, this transform can be applied to the 3D coordinates of the LEDs obtained in step 2, in order to determine their position in the arena frame:

$$(x_A, y_A, z_A) = T_{A \rightarrow C}^{-1} (x_c, y_c, z_c) \quad (14)$$

Where $T_{A \rightarrow C}$ is the transformation matrix defined by the basis vectors \mathbf{X}_C , \mathbf{Y}_C and \mathbf{Z}_C . Note, that the transformations $T_{A \rightarrow C}$ and $T_{C \rightarrow A}^{-1}$ include both translation and rotation of the coordinate frame.

Step 4: in step 1 we used the fact that all three Euler angles with respect to a given coordinate frame can be determined from just the x-y projection of the LED-coordinates. Hence, it is sufficient to consider the x-y values of the reconstructed device in the arena frame, and then to re-apply the relations used in step 1, in order to obtain the final angles with respect to the arena. Namely from equations (4), (9) and (11) we can derive the Euler angles in arena coordinates:

$$\gamma_A = \arctan [(\alpha_0 - \alpha_A) (r_1 + r_2) / h] \quad (15)$$

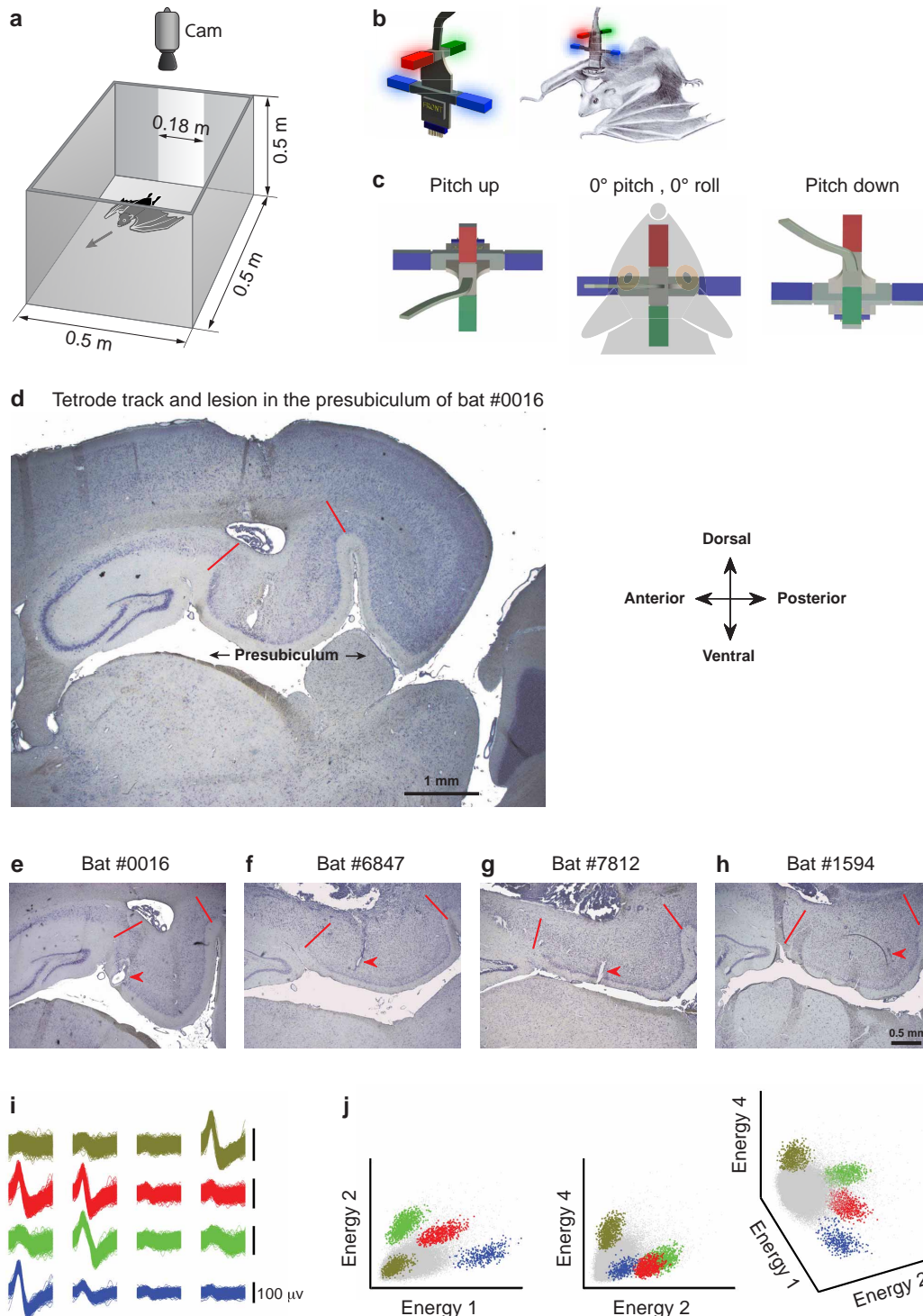
$$\theta_A = \arctan [(\beta_0 - \beta_A) (f_1 + f_2) \cos(\gamma_A) / h] \quad (16)$$

$$\varphi_A = \text{atan}[v_y / v_x] + \pi H(-v_x) \quad (17)$$

Where \mathbf{v} is defined similarly to equation (10), but for the projection onto the arena plane.

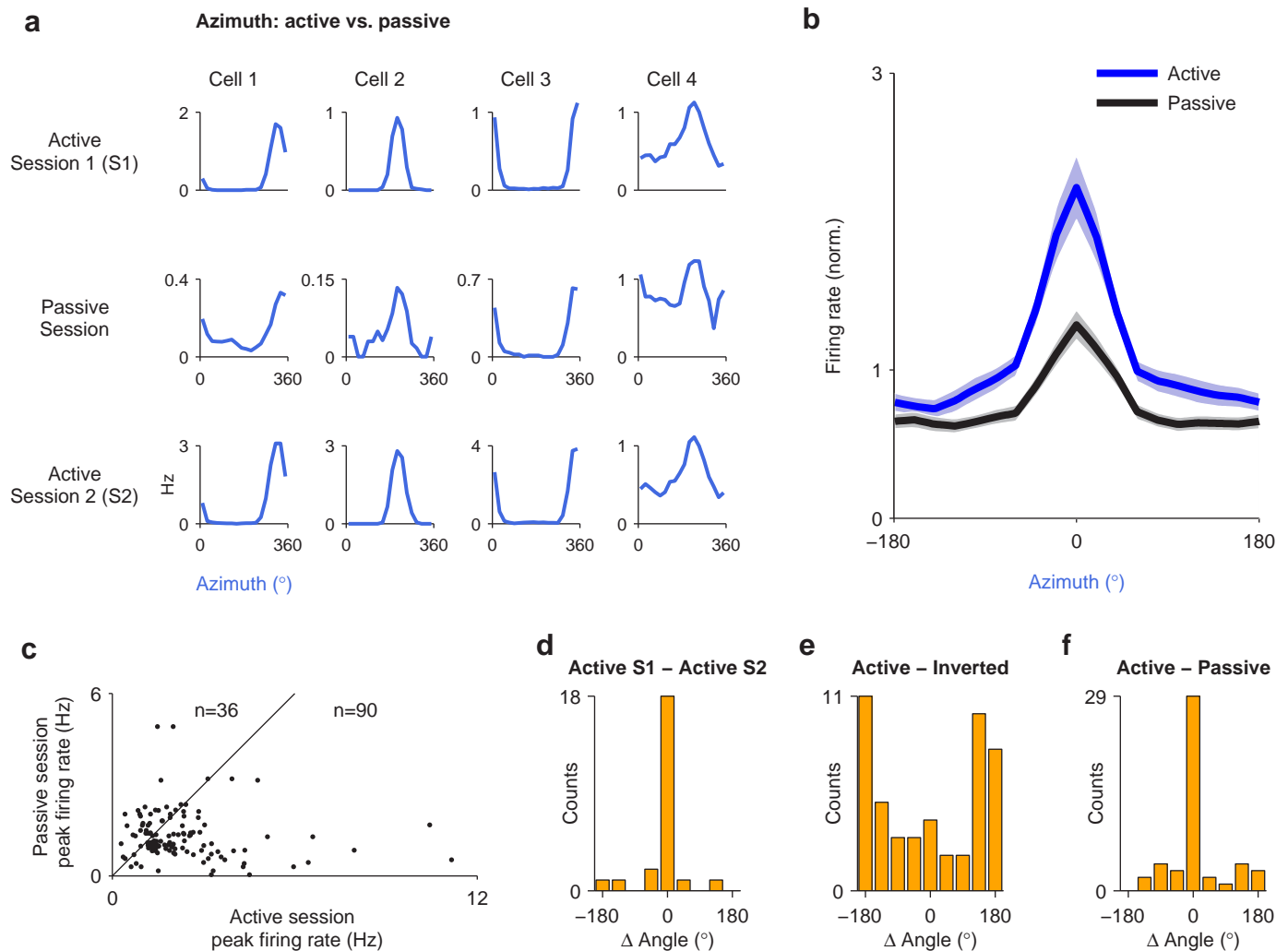
The Euler angles (φ_A , θ_A , γ_A) in the arena coordinates, which we computed here (equations 15–17), were the head-direction angles (φ , azimuth; θ , pitch; γ , roll) that were reported in this study.

51. Neuweiler, G. *The Biology of Bats* (Oxford Univ. Press, 2000).
52. Altringham, J. D. *Bats: Biology and Behaviour* (Oxford Univ. Press, 1996).
53. Yovel, Y., Geva-Sagiv, M. & Ulanovsky, N. Click-based echolocation in bats: not so primitive after all. *J. Comp. Physiol. A* **197**, 515–530 (2011).
54. Ulanovsky, N. & Moss, C. F. Dynamics of hippocampal spatial representation in echolocating bats. *Hippocampus* **21**, 150–161 (2011).
55. Zar, J. H. *Biostatistical Analysis* 4th edn (Prentice Hall, 1998).
56. Skaggs, W. E., McNaughton, B. L., Gothard, K. M. & Markus, E. J. An Information-theoretic approach to deciphering the hippocampal code. *Advances in Neural Information Processing Systems* **5**, 1030–1037 (1993).
57. Skaggs, W. E., McNaughton, B. L., Wilson, M. A. & Barnes, C. A. Theta phase precession in hippocampal neuronal populations and the compression of temporal sequences. *Hippocampus* **6**, 149–172 (1996).
58. Cacucci, F., Lever, C., Wills, T. J., Burgess, N. & O'Keefe, J. Theta-modulated place-by-direction cells in the hippocampal formation in the rat. *J. Neurosci.* **24**, 8265–8277 (2004).
59. Sharp, P. E. Multiple spatial/behavioral correlates for cells in the rat postsubiculum: multiple regression analysis and comparison to other hippocampal areas. *Cereb. Cortex* **6**, 238–259 (1996).



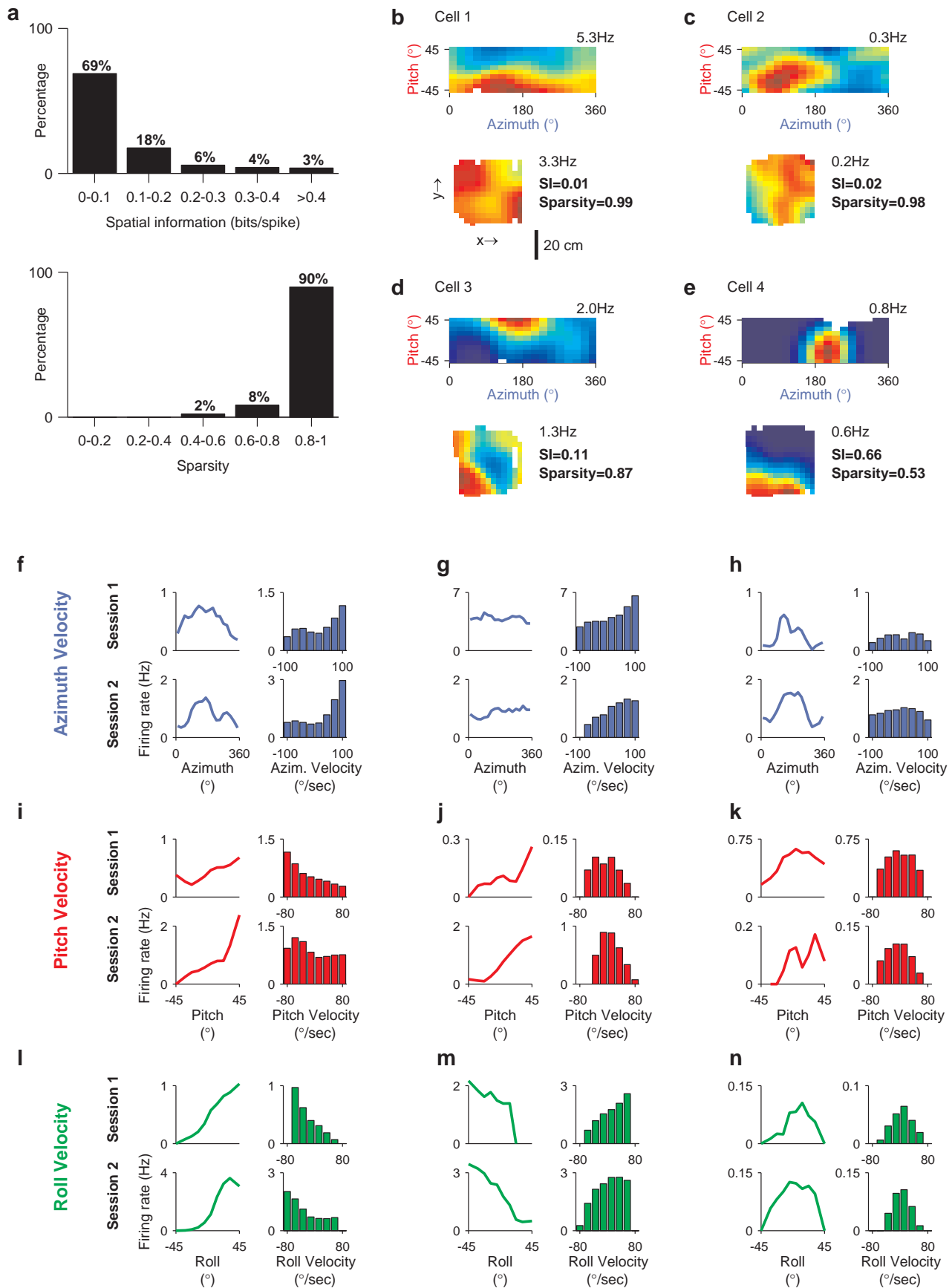
Extended Data Figure 1 | Experimental methods for behavioural setup number 1, recording locations, and example of spike-sorting. **a**, Schematic illustration of the behavioural arena and camera position used in the crawling experiments (setup number 1). **b**, Illustrations of the device used for computing the head-direction Euler angles, which was based on a 3D non-coplanar arrangement of four LEDs. This four-LED device was mounted on the recording headstage, and allowed measuring the Euler angles using one overhead camera (see Methods and Extended Data Figs 11 and 12 for details of the algorithm). **c**, Schematic camera views of the four-LED headstage at different pitch angles. **d**, Nissl-stained sagittal brain section through the hippocampal formation of an Egyptian fruit bat, including the dorsal presubiculum (same brain section as in Fig. 1c). Electrolytic lesions were done at the end of the experiment. Red lines denote the borders of the presubiculum. Scale bar, 1 mm. **e–h**, Sagittal Nissl-stained brain sections showing

representative recording sites in the presubiculum of 4 bats. Tetrode tracks are marked by red arrowheads (the arrowheads point to lesion sites in bats number 0016 and 6847; while in bats number 7812 and 1594, the arrowheads point to the end and to the middle of the tetrode-track, respectively). Sections in **d** and **e** are from the same animal (bat 0016); the section in **e** shows the track of a different tetrode than the one seen in **d**. Scale bar in **e–h**, 0.5 mm. **i**, Waveforms of four different neurons (different colours, rows) recorded simultaneously on the 4 channels of a single tetrode (columns). Scale bar, 100 μ V; waveform duration, 1 ms. **j**, Energy displays (cluster plots) for the data in **i**, showing the energy of spikes (dots) on two of the tetrode's four channels (left and middle panels), or on three of the tetrode's four channels (right panel). Colours match the waveforms in **i**; grey dots, small spikes or noise that crossed the voltage threshold but were not classified as single units.



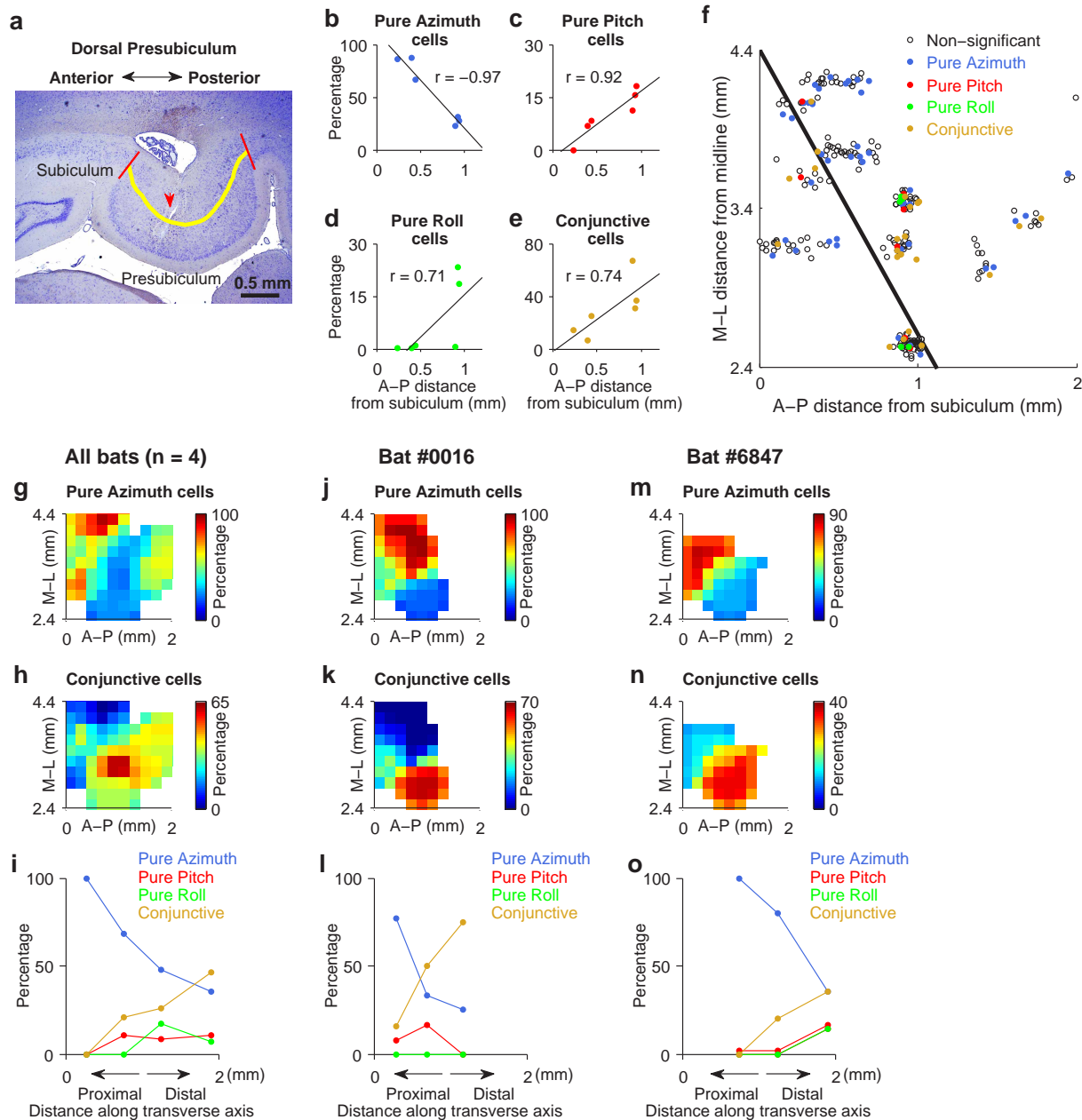
Extended Data Figure 2 | Head-direction cells maintain their preferred azimuthal direction when the bat is being moved passively in the upright position. **a**, Azimuth tuning of four example cells in three upright sessions: active session number 1 (top row), passive session (middle), and active session number 2 (bottom). Note that the preferred direction in the passive session stayed similar to the preferred direction of the active sessions, despite the reduction in cells' firing rate during the passive session. **b**, Population average tuning curve for active (blue) versus passive (black) sessions, for all azimuth-significant cells ($n = 63$ cells); shading, mean \pm s.e.m. Prior to averaging, the tuning-curve of each neuron was centred around its peak, and the firing rate was normalized to the mean firing rate of the neuron across all sessions (active and passive pooled together). Note that, on average, presubiculum neurons exhibited a lower firing rate during the passive session as compared to the active sessions. **c**, Peak firing rate of azimuth-tuned neurons ($n = 63$ cells, that is, 126 sessions) in the active versus passive sessions, showing a significantly lower firing rate during the passive condition (sign test, $P < 0.001$; number of sessions above and below the diagonal are indicated on the graph). The peak firing rates of these neurons ranged up to 12 Hz, and were generally lower than peak

firing rates in rats, similar to what we found previously for hippocampal place-cells and entorhinal grid-cells in crawling bats^{9,16,54}, and consistent with the slow crawling velocity of bats, which reduces the firing rate^{36,40}. Importantly, even cells with low firing rates exhibited stable directional tuning across sessions (see panel **a**). **d–f**, Changes in tuning properties during inversion are not caused by the passivity of the movement in the inverted session. **d**, Distribution of angular differences in the preferred azimuthal head direction between the two active behavioural sessions (S1 and S2). The histogram was plotted for cells with significant tuning in the upright active sessions that also had significant tuning in the inverted session. Peak around 0° indicates stability of the azimuthal tuning across the two active behavioural sessions. **e**, Angular difference between the upright sessions and the inverted session. Peaks at around $\pm 180^\circ$ indicate that inversion of the bat upside-down resulted in a 180° shift in the preferred azimuthal direction for the azimuth cells. **f**, Angular difference between the active upright sessions and the passive (upright) session. Peak around 0° indicates that passive movement in itself does not induce a change in the preferred direction of the neurons (see also examples in **a**).



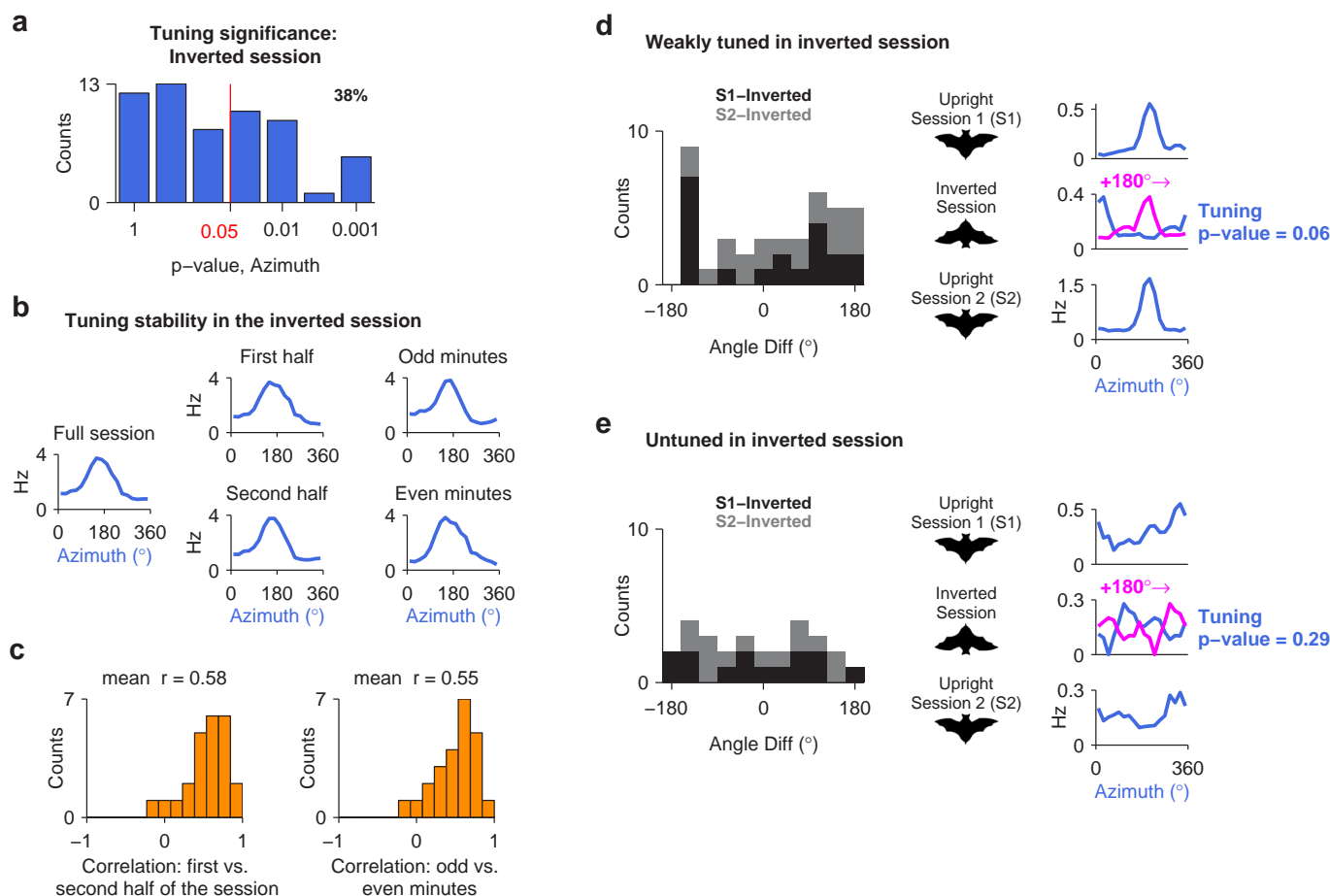
Extended Data Figure 3 | Bat presubiculum neurons are modulated by angular velocity but exhibit weak spatial tuning. **a**, Distribution of spatial information (top) and sparsity (bottom) for all the 122 presubiculum neurons recorded in setup number 1. Note that, with a few exceptions, neurons in the presubiculum conveyed little information about the location of the bat in the environment and displayed weak spatial selectivity, as indicated by their low spatial-information index and high sparsity index. **b–e**, Examples of four neurons, depicting the head-direction fields (top) and place-fields (bottom); spatial information (SI) and sparsity for each neuron are indicated next to the positional plot; the cells were ordered according to spatial information values. Most neurons (**b–d**) carried little spatial information and showed irregular spatial firing patterns, despite having significant directional responses. Only 3% of the neurons had spatial information > 0.4 bits per spike (for example, the cell in **e**), in addition to their directional tuning. This small minority of cells could possibly be border/boundary cells^{18–20}, or place-by-direction cells^{37,58}, or perhaps conjunctive grid \times head-direction cells^{15,24} (we could not detect grid structure because of the small size of the arena, 50×50 cm). **f–n**, Head-direction cell activity is modulated by angular velocity. Shown is the tuning to head direction (left) and angular velocity (right) for 9 examples from 8 neurons

(panels **f** and **i** are from the same neuron). Top row (**f–h**): azimuth and azimuth-velocity; middle row (**i–k**): pitch and pitch-velocity; bottom row (**l–n**): roll and roll-velocity. **f**, A neuron tuned to azimuth (left panels) that increases its firing rate in response to faster turning in the clockwise direction (right panels). **g**, A neuron that did not exhibit any tuning to head direction in azimuth (left panels), but nevertheless increased its firing rate with increasing head velocity in the clockwise direction (right panels); this neuron might be classified as a pure ‘angular velocity neuron’. **h**, A neuron tuned to azimuth that exhibited almost no modulation by azimuth-velocity; this neuron might be classified as a pure ‘head-direction neuron’. **i**, A neuron tuned to pitch, which increased its firing rate in response to fast decrease in pitch. Note that this is in fact a conjunctive azimuth \times pitch neuron (same neuron as in **f**). **j–k**, Pitch-tuned neurons that also fired preferentially to slow angular head-velocity in pitch. **l–n**, Roll-neurons that were also modulated by anti-clockwise angular-velocity in roll (**l**), clockwise angular-velocity in roll (**m**), or slow-roll velocity (**n**). These diverse types of tuning to angular-velocity resemble the functional diversity of azimuth-velocity tuning reported for presubiculum neurons in the rat^{4,59}, and also suggest that bat presubiculum neurons can integrate head angular-movements in rotation planes other than Earth’s horizontal plane.



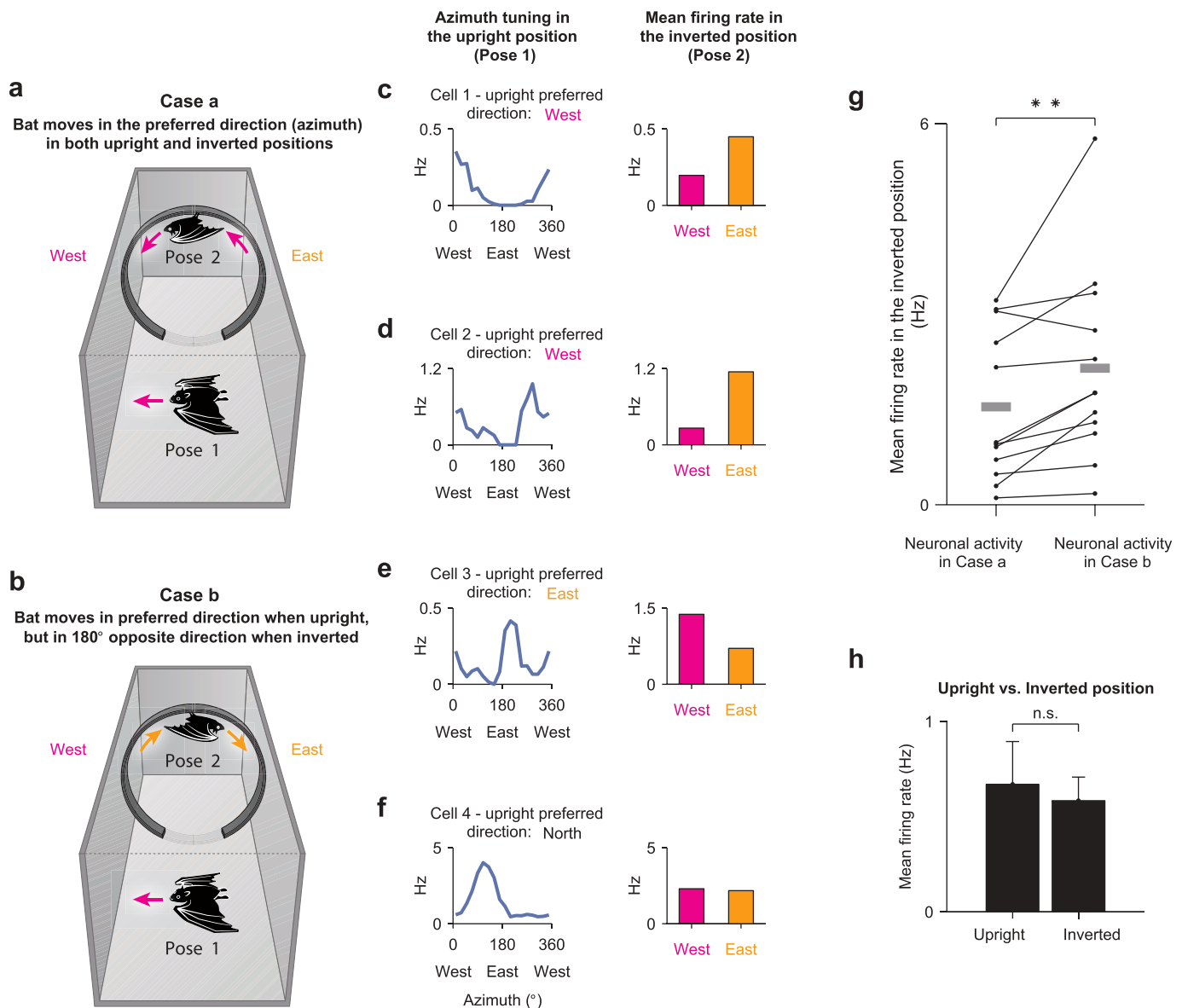
Extended Data Figure 4 | Functional-anatomical gradient of head-direction cells along the transverse axis of dorsal presubiculum. **a**, Sagittal section through the presubiculum; red lines, presubiculum borders. Yellow line shows the axis used to determine the anterior-posterior distance of the tetrode-track (arrowhead) from the subiculum border. Scale bar, 0.5 mm. **b–e**, Percentage of head-direction cells in each tetrode-track, plotted against the anterior-posterior ('A-P') distance of that tetrode-track from the subiculum. Each dot represents one tetrode-track; shown are only tetrode-tracks with > 5 significant head-direction cells recorded per track. Percentages plotted separately for neurons with pure tuning to azimuth (**b**), pitch (**c**), or roll (**d**), and for neurons with conjunctive tuning to any angular combination (**e**); regression lines are also indicated. **f**, 2D unfolded map of dorsal presubiculum, showing the reconstructed recording location for all the 266 neurons recorded from the 4 bats in setup number 1. Colours indicate pure tuning to azimuth (blue), pitch (red), or roll (green), or neurons with conjunctive tuning to any angular combination (yellow); cells that did not pass the criterion for stability or directionality are also shown (empty circles). Note that the antero-posterior (A-P) distance is measured along the curved structure of the presubiculum (as shown by the yellow line in Fig. 2h), while the medio-lateral (M-L) distance is measured straight from midline. The multilinear regression line (black) was computed from the 2D maps in **g**, **h**. Dots were slightly jittered, for display purposes only, to prevent overlap. **g**, **h**, Two-dimensional unfolded maps of

dorsal presubiculum from the 4 bats tested in setup number 1, showing the percentage of pure azimuth (**g**) and conjunctive cells (**h**), based on the recording locations of individual cells (M-L, medio-lateral distance from midline; Methods). These maps reveal a functional gradient along a diagonal axis, with pure azimuth cells dominating the anterolateral part and conjunctive cells the posteromedial part of dorsal presubiculum. From these maps, we calculated the multilinear regression line (black line in **f**), based on A-P and M-L positions as predictors. This line represents the axis of the maximal diagonal gradient of head-direction cells, and is computed as the slope of the multilinear regression coefficients, averaged for pure azimuth and for conjunctive cells, which together represent the majority of the head-direction cells. **i**, Percentages of all 4 neuronal types from **b–e**, binned here along the diagonal axis (regression line) of the functional gradient, which corresponds to the transverse axis of the presubiculum (Methods; graph based on all the significant head-direction cells from the 4 bats in setup number 1: $n = 78$ neurons). **j–o**, Same as in **g–i**, computed now for two individual bats in which at least 3 tetrodes were identified within the dorsal presubiculum, which is the minimum number of tetrode-tracks that allowed us to create a functional-anatomical map for an individual animal (**j–l**, bat number 0016, $n = 36$ significant head-direction cells; **m–o**, bat number 6847, $n = 21$ significant head-direction cells).



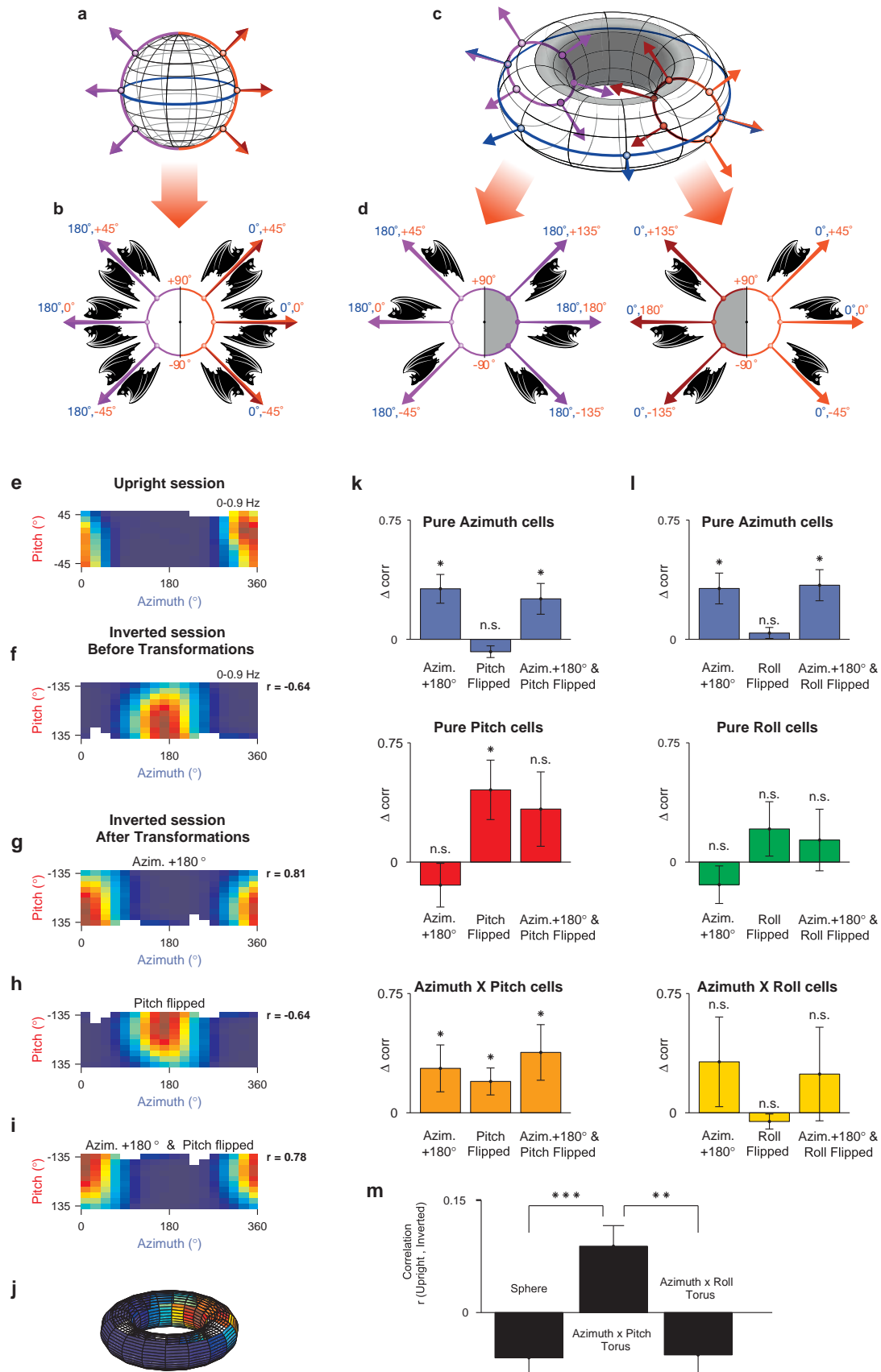
Extended Data Figure 5 | Azimuth tuning and stability analysis for neurons in the inverted session. **a**, Distribution of tuning significance of azimuth-encoding cells ($n = 63$), in the inverted session (setup number 1). Red line, 95th percentile value from shuffled data. **b**, **c**, Stability analysis of azimuth cells in the inverted session. **b**, Example of azimuth tuning for a single cell. Left, tuning curve computed for the entire inverted session. Middle column, comparing the tuning curve computed for the first half versus the second half of the inverted session. Right column, comparing the tuning curve computed for odd versus even minutes of the inverted session. **c**, Distribution of correlation coefficients for azimuth cells with significant tuning in the inverted session ($n = 24$), parsed based on two partitioning conditions. Left, first half versus second half of the session; right, odd versus even minutes. **d**, Left, distributions of differences in preferred direction between sessions, for cells that have not reached tuning-significance criterion (that is, neurons to the left of the red line in **a**), but were nevertheless somewhat tuned in azimuth in the inverted session ('weakly tuned', $n = 20$; see Methods). Distribution of angular differences

between each of the upright sessions and the inverted session (S1-inverted, S2-inverted) of these 'weakly tuned' neurons shows peaks near $\pm 180^\circ$ for the inverted session, indicating 180° -shift of the preferred direction, similar to that observed in azimuth cells that were significantly tuned under inversion (Fig. 3b, bottom). Right, azimuth tuning of an example cell that did not pass the shuffling criterion for tuning under inversion, but was close to significance ($P = 0.06$); this neuron was considered 'weakly tuned' by our criteria, but nevertheless this neuron showed a shift of 180° in the inverted session (middle plot, blue) relative to the upright sessions (top and bottom). Magenta curve in the middle row: the tuning-curve of the inverted session, shifted by 180° . **e**, Left, same as in **d**, for the remainder of the cells that have not reached significance criterion and were indeed untuned in the inverted session (that is, these cells were to the left of the red line in panel **a**, and were also very clearly not directionally tuned: 'untuned cells', $n = 11$; Methods). Right, an example cell with very little tuning under inversion (test for tuning significance for this neuron: $P = 0.29$).



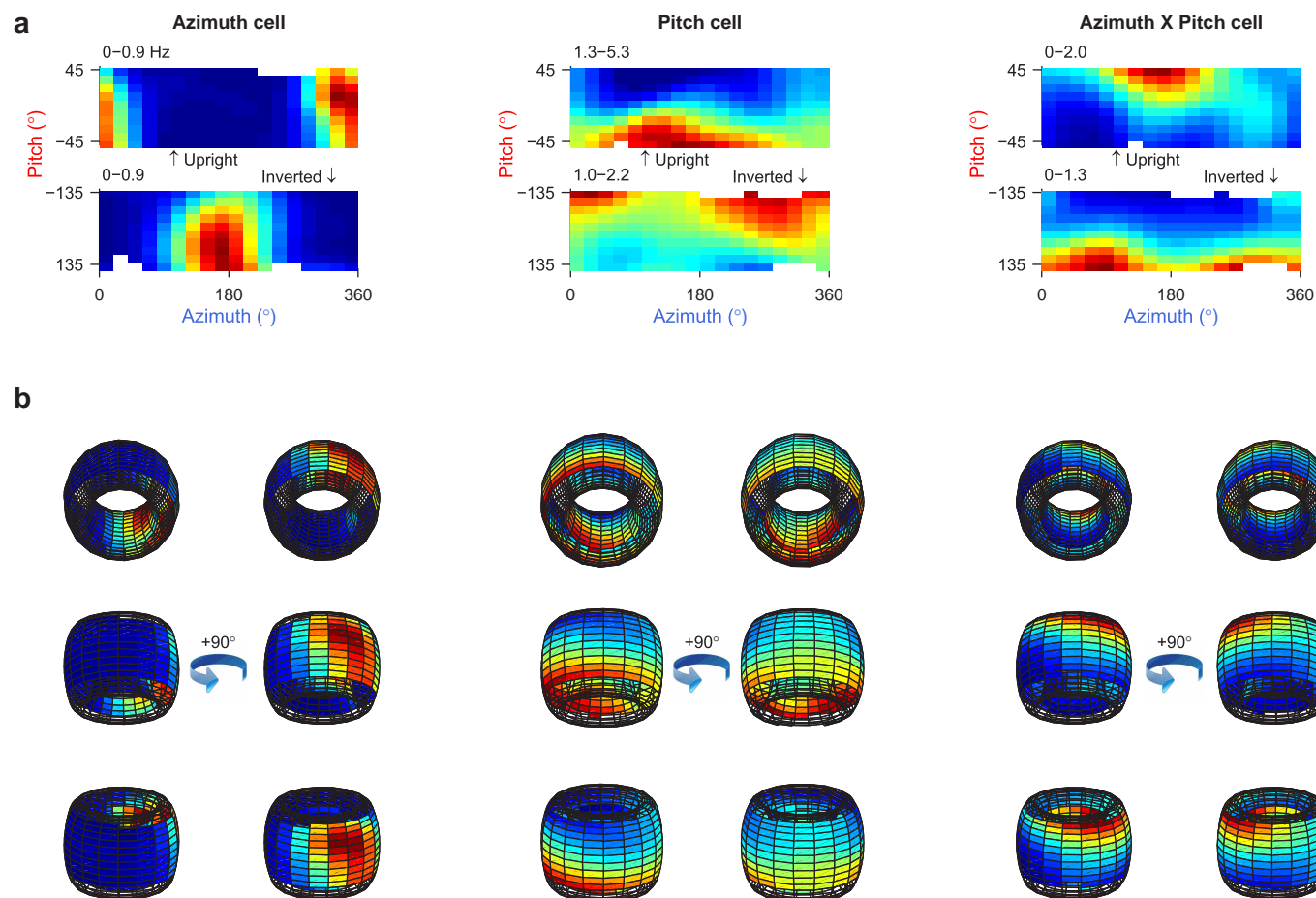
Extended Data Figure 6 | Azimuth cells exhibit a 180°-shift in their preferred direction when bats position themselves upside-down on their own volition. **a, b**, Schematic illustration of the vertical-ring apparatus (setup number 2), designed to compare the tuning properties of neurons during upright crawling on the arena floor (pose 1) with the activity of the same neurons during epochs in which the bat had positioned itself upside-down on its own volition along the inner side of a vertical ring (pose 2). The analysis was restricted to azimuth cells whose preferred direction on the arena floor was aligned to the cardinal axis of the ring (west–east), so that movement of the inverted bat would be either in the upright-preferred-direction of the neuron (case a), or in the exact opposite (180°) direction (case b). **c–e**, Examples of azimuth cells, showing reversed firing in animals that positioned themselves upside down on their own volition. Left, azimuth tuning curves of individuals neurons during upright locomotion on the arena floor (pose 1). Right, mean firing rates of the same neurons during active inverted motion (pose 2), shown separately for west and east directions. **c, d**, Two example cells with west direction tuning on the arena floor (left, blue curves), showing a reversal of their

firing direction in the inverted position (right, increase in firing to the east). **e**, Example cell with east direction tuning on the arena floor, showing a reversal of its firing direction in the inverted position (increase in firing to the west). **f**, Example cell, whose preferred direction on the floor was orthogonal to the cardinal axis of the ring (preferred direction was to the north). Note that this neuron showed no preference to either west or east directions in the inverted position (right). **g**, Pairwise comparison of mean firing rates of azimuth cells in the inverted position, for epochs in which the inverted bat moved in the upright-preferred-direction of the neuron (case a, as shown in panel **a**), versus epochs in which it moved in the opposite (180°) direction (case b, as shown in panel **b**). Notably, 92% of the cells ($n = 11/12$) increased their firing rate when the head-azimuth in the inverted position was 180° opposite to the upright-preferred-direction of the cells, as predicted by the torus model. $^{***}P < 0.01$. **h**, Comparison of mean firing rates in the upright position versus the inverted position for the azimuth cells analysed in **g**, showing no significant change in the mean firing rate under inversion. Error bars, mean \pm s.e.m.



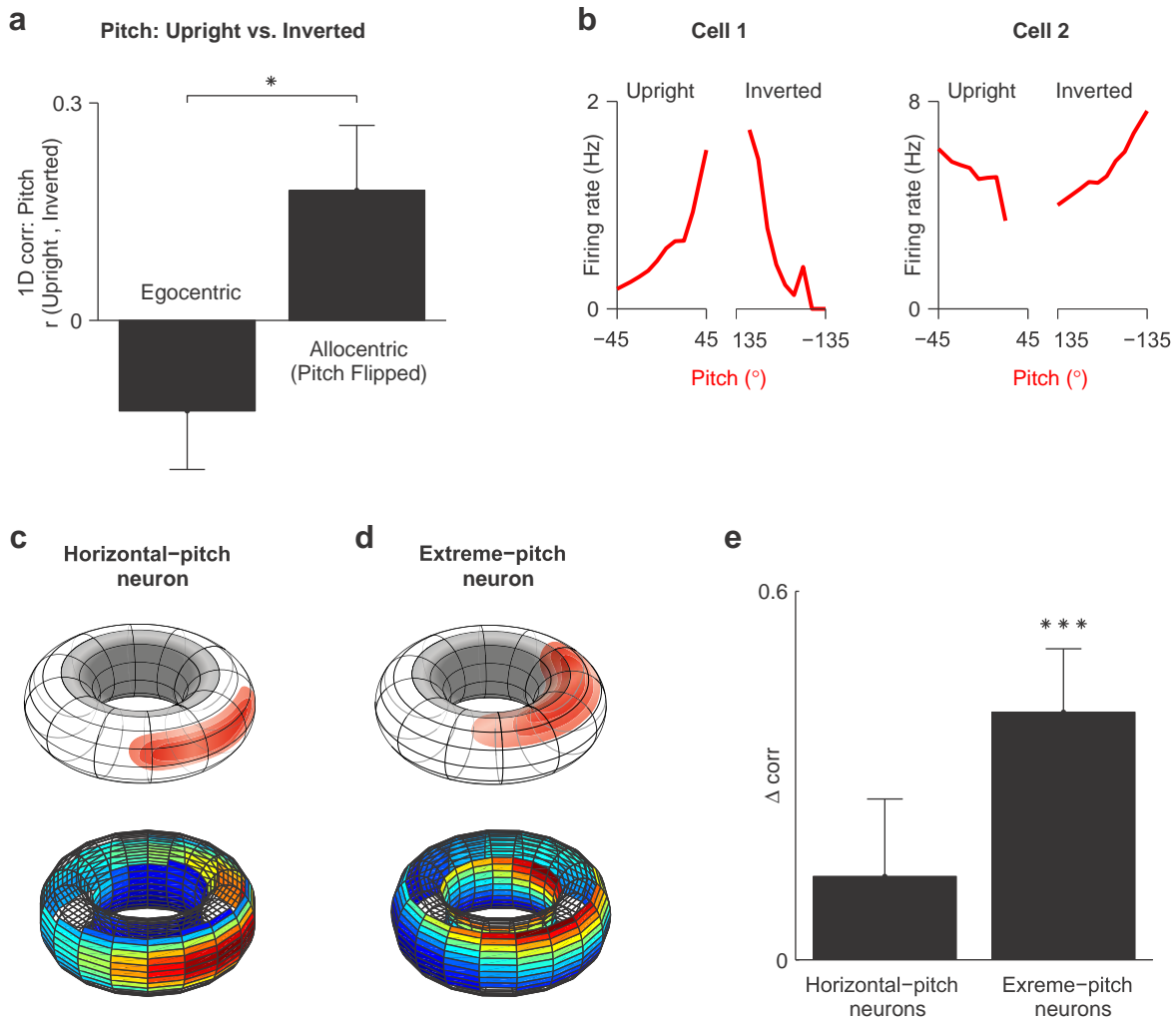
Extended Data Figure 7 | Spherical versus toroidal coordinate systems. **a**, In a spherical coordinate system, which describes the direction of a vector in 3D space, 3D head direction is defined by its azimuth and pitch angles. Each line of longitude on a sphere corresponds to a specific azimuth. The position of the head along this line of longitude is given by its pitch angle (ranging from -90° to $+90^\circ$). For example, the east longitude (red-coloured half-ring) corresponds to different pitch angles along the 0° azimuth. The west longitude (purple-coloured half-ring) corresponds to different pitch angles along the 180° azimuth. **b**, The azimuth and pitch angles of the head in spherical coordinates define its absolute direction in 3D, regardless of whether the bat is upright or inverted. For example, the spherical coordinates of the bat furthest on the right in this plot (0° azimuth, 0° pitch), correspond to an upright bat with its head parallel to the horizontal plane facing the east direction but also to an inverted bat facing east (also parallel to the ground). Thus, the spherical representation is ambiguous with respect to the upright versus the inverted state. **c**, In the toroidal coordinate system, the two angles that determine the direction of the animal's head in 3D (azimuth and pitch) represent two independent cyclic degrees of freedom, both having a range of 360° (blue circle, azimuth; red and purple circles, pitch). Importantly, the toroidal azimuth does not represent the direction of the animal's nose (rostral-caudal axis) relative to a distal point in space, but rather the direction of the inter-aural axis. The toroidal pitch, in contrast, is anchored to the direction of the rostral-caudal axis. In this representation, any rotation in pitch does not change the azimuth, as defined in toroidal coordinates (Methods). Specifically, if the bat pitches its head to angles greater than $+90^\circ$ pitch (resulting in flipping from upright to inverted position), the toroidal azimuth still remains the same. Importantly, the upright and inverted positions are represented continuously, but can be distinguished according to the pitch angle: The outer surface of the torus (white) corresponds to all upright positions ($-90^\circ < \text{pitch} < +90^\circ$), whereas the inner part of the torus (grey) corresponds to all inverted positions ($+90^\circ < \text{pitch} < +180^\circ$ or $-180^\circ < \text{pitch} < -90^\circ$). **d**, Detailed depiction of two different azimuthal directions on the torus (shown as red and purple rings in **c**). Right panel (0° azimuth): for an upright bat facing east (0° azimuth, 0° pitch), any change of the pitch angle (red ring) will not change the toroidal azimuth, which will remain 0° . Left panel (180° azimuth): analogously, for an upright bat facing west (180° azimuth, 0° pitch), any change of the pitch angle (purple ring) will not change the toroidal azimuth, which will stay 180° . Note, that this set of positions, corresponding to 180° azimuth (purple ring), is mapped onto the opposite side of the torus, relative to the set of positions corresponding to 0° azimuth (red ring). Therefore, unlike the ambiguous representation of head direction in spherical coordinates (**a**, **b**), there is no ambiguity in the toroidal representation: each point on the torus describes a unique orientation of the bat, and defines not only its head direction in 3D, but

also whether the bat is in the upright or in the inverted position. **e–j**, Example of construction of the toroidal representation and angular transformations of the 2D rate-maps for the inverted session, for one pure azimuth neuron. **e**, **f**, 2D directional rate maps for the upright session (**e**) and inverted session (**f**) computed initially in spherical coordinates. **g–i**, Angular transformations of the 2D rate map of the same inverted session (shown in **f**), done in order to represent it in toroidal coordinates: **g**, 180° shift in azimuth; **h**, flipping the pitch axis (to represent it continuously and allocentrically); **i**, both transformations together (that is, 180° shift in azimuth and flipping the pitch axis). For each of the maps in **f–i**, the value of the 2D Pearson correlation coefficient, r , of this map with the upright map in **e**, is also indicated. **j**, The toroidal representation is constructed by concatenating the 2D directional map of the upright session (shown in **e**) with the map of the inverted session after both transformations (shown in **i**). **k**, Difference in 2D correlations (Δcorr) of the firing maps in the upright and inverted sessions, before (**f**) versus after the various angular transformations (examples shown in **g–i**), which included shifting the azimuth of the inverted session by 180° , or flipping the pitch in order to represent it allocentrically, or both. Error bars, mean \pm s.e.m.; $*P < 0.05$. Comparisons are shown for pure azimuth cells (**k**, upper panel, $n = 84$ (42 cells \times 2 sessions)), pure pitch cells (middle, $n = 14$), and conjunctive azimuth \times pitch cells (lower panel, $n = 28$). Consistent with the toroidal model, a 180° azimuth-shift of the inverted map increased substantially the correlation with the upright-map for pure azimuth cells (**k**, upper panel-left bar; t -test, $P < 0.05$), but not for pure pitch cells (middle panel left bar: t -test, NS). Conversely, flipping the pitch increased substantially the correlation for pure pitch cells (middle panel middle bar; t -test, $P < 0.05$), but not for pure azimuth cells (upper panel middle bar: t -test, NS). For the azimuth \times pitch conjunctive cells, both shifting the azimuth by 180° and flipping the pitch resulted in significant increase in correlation values (lower panel; t -test: $P < 0.05$, for all bars), as expected from cells encoding both azimuth and pitch. **l**, Similar analysis for azimuth \times roll 2D maps did not show a significant effect of roll on the correlations between the upright and the inverted sessions, suggesting that the roll dimension is not crucial for 3D head-direction representation in bats. **m**, Population averages of 2D correlations of the firing maps in the upright versus inverted session, for all the head-direction cells, including azimuth, pitch and roll cells ($n = 156$ (78 cells \times 2 sessions)), when represented in either spherical coordinates (left), or azimuth \times pitch toroidal coordinates (middle), or azimuth \times roll toroidal coordinates (right). These results suggest that an azimuth \times pitch torus, but not the alternative models such as the sphere or an azimuth \times roll torus, captures well the activity of head-direction cells in the bat presubiculum. Error bars, mean \pm s.e.m.; $**P < 0.01$; $***P < 0.001$.



Extended Data Figure 8 | Toroidal representation of head-direction cells. **a**, 2D directional rate-maps of the upright (top) and inverted session (bottom), for a pure azimuth cell (left), pure pitch cell (middle), and a conjunctive azimuth \times pitch cell (right); same neurons as in Fig. 3f of the main text. **b**, Same three cells as in **a**, shown in the toroidal representation from different viewing angles. Each cell is shown from two horizontal viewing angles (rotated 90°

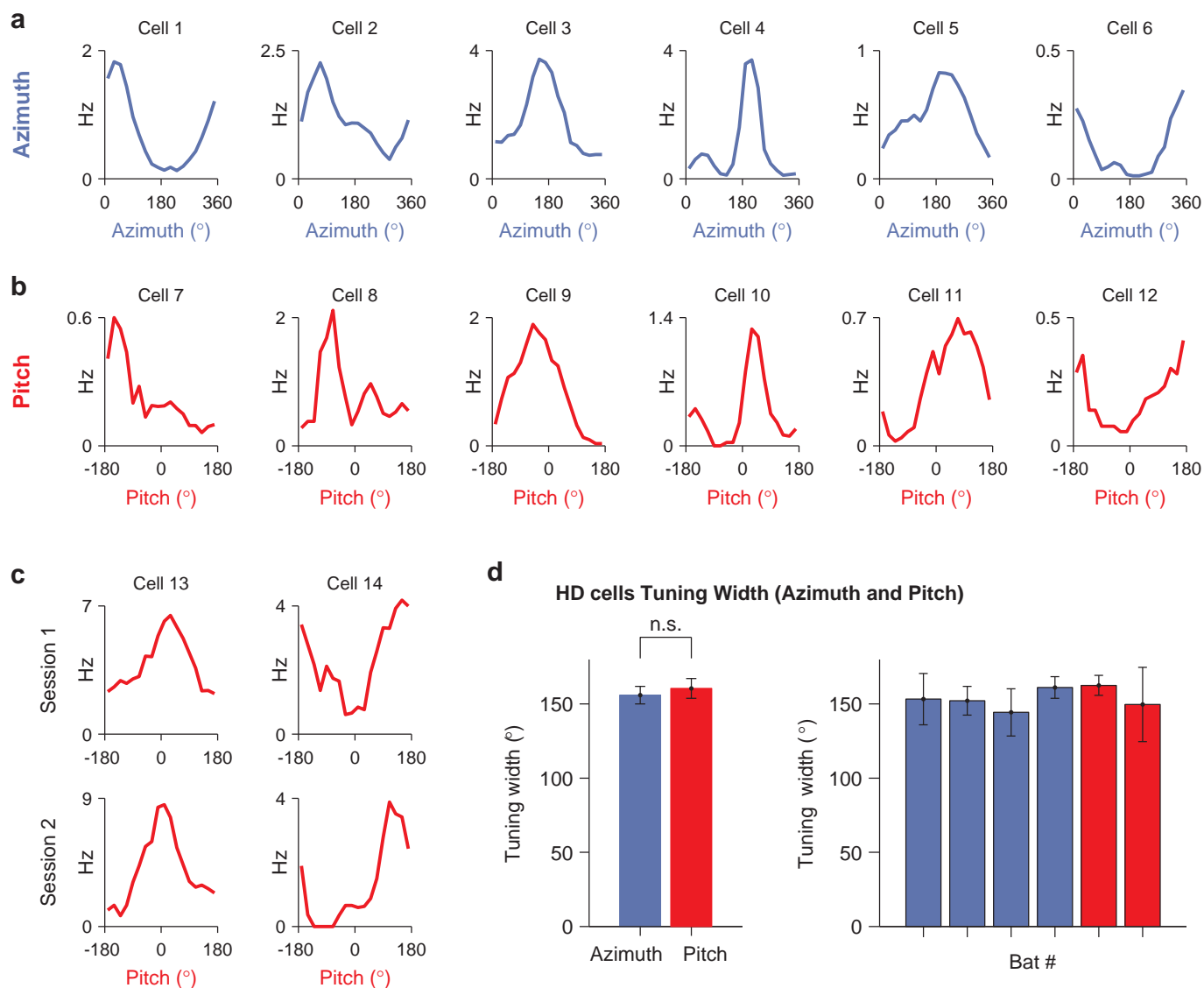
horizontally (azimuthally) with respect to each other) and from three different pitch viewing angles. The tori were constructed by plotting on the outside half of the torus the 2D directional rate map for the upright session, and on the inside half of the torus plotting the rate map of the inverted session in toroidal coordinates; see Extended Data Fig. 7e–i and Methods for the details of the angular transformations.



Extended Data Figure 9 | Torus topology predicts that tuning to pitch is allocentric and distinct between upright and inverted positions.

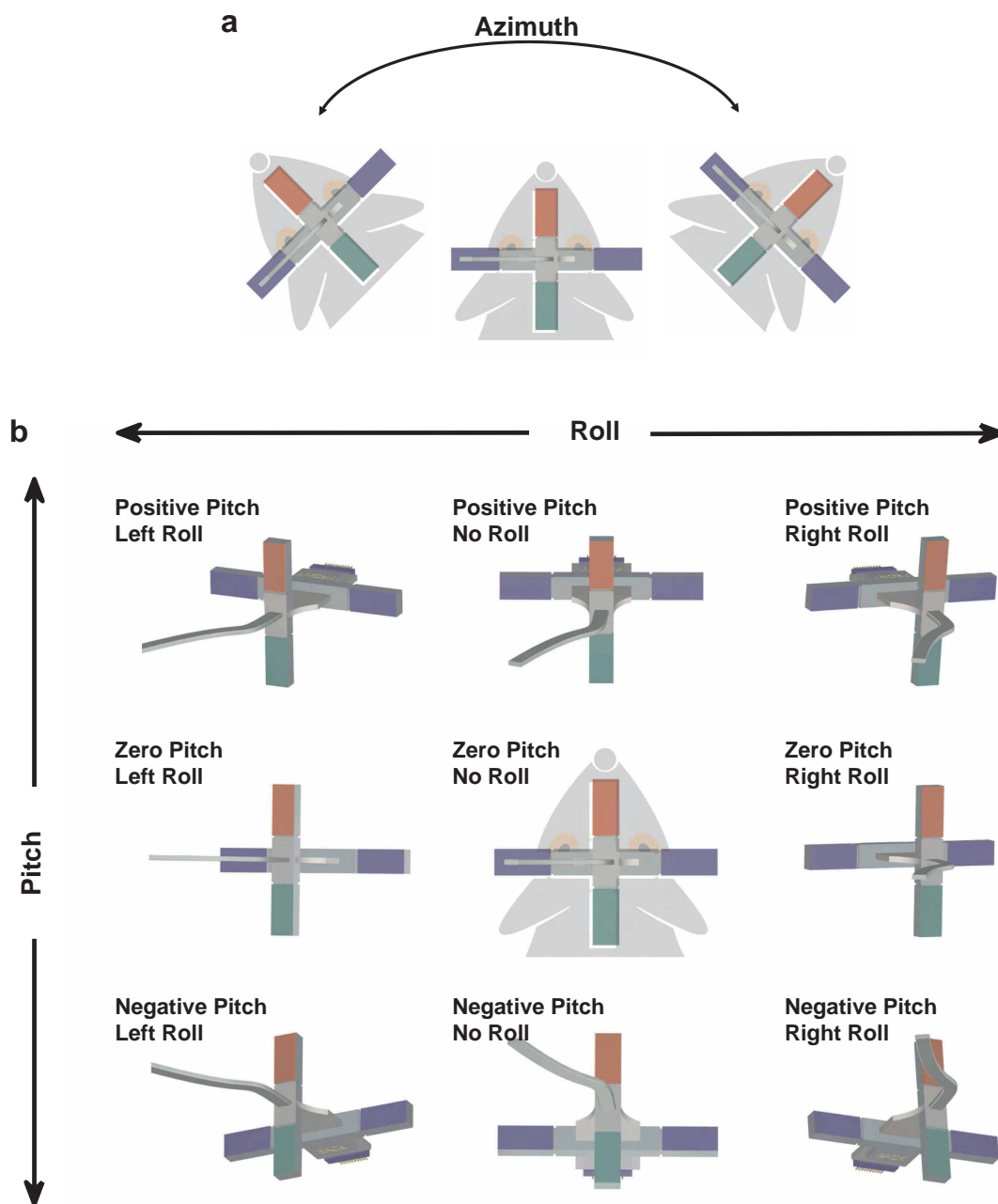
a, According to the toroidal representation, pitch is computed in a world reference frame (allocentric) and not in body reference frame (egocentric). In the upright position, the two reference frames are indistinguishable. For example, when a bat pitches its head towards the moon (positive allocentric pitch) it also raises its head away from its chest (positive egocentric pitch). However, in the inverted position, allocentric head pitch is flipped with respect to the egocentric one. When the bat is upside-down and looks towards the moon (positive allocentric pitch), it now brings the head towards the chest (negative egocentric pitch). To test which of these reference frames is most consistent with our neural data, we computed the correlation between the pitch 1D tuning curves of the upright sessions versus the inverted session, in the two reference frames (for the experiments in setup number 1). Correlation of pitch tuning-curves between the upright and inverted positions was higher when the inverted session was plotted in allocentric coordinates ($n = 42$ (21 cells \times 2) upright sessions; we included in the analysis only cells that were significantly tuned to pitch). $*P < 0.05$. **b**, A toroidal representation implies that pitch has a continuous representation, where every pitch angle corresponds to a unique orientation along a 360° ring of possible pitch angles (see Fig. 3d, red ring). This implies that if a neuron is active mostly at extreme pitch angles during the upright session ('extreme-pitch' neuron), it is likely to be active also at the contiguous pitch in the inverted session. Shown here are examples of two pitch cells with tuning to non-zero pitch angles in the upright session (cell 1, positive pitch; cell 2, negative pitch). 1D tuning to pitch is plotted for the average neuronal activity of the cell during the two upright sessions ('upright', left), and for the inverted session ('inverted', right). Note that the two cells exhibit contiguous firing in the inverted and upright sessions. **c–e**, The toroidal model generates a prediction, that such a continuity between the upright and the inverted session (as shown in **b**), should occur for cells tuned to 'extreme pitch' (see example in **d**), but not for cells tuned to horizontal pitch (example in **c**). More specifically, in the toroidal model, neurons with preferred pitch at

around 0° , an angle at which the head of an upright bat is parallel to the ground ('horizontal pitch' cells), are not expected to fire when the bat is inverted with its head being parallel to the ground, because these two situations are topologically distinct in the toroidal but not in the spherical representation (Fig. 3d vs 3c and Extended Data Fig. 7d vs 7b). In contrast, neurons tuned to an extreme pitch angle in the upright position, are likely to fire to some extent also in the contiguous part of the inverted session, as the 'patch of activity' on the 'external side' of the torus (which corresponds to upright position) is likely to extend also onto the 'inner side' of the torus (corresponding to inverted position). Therefore, according to the toroidal (but not the spherical) model, the correlations between the upright and inverted sessions for cells tuned to 'extreme pitch' are expected to be higher than for cells tuned to 'horizontal pitch'. This prediction was tested here, and was indeed confirmed (see below). **c**, Upper panel, schematic representation of an azimuth \times pitch cell, exhibiting pitch tuning to 0° (a 'horizontal pitch' neuron). Lower panel, example of an actual neuron exhibiting pitch tuning to 0° , similar to the schematic. Note that in both the schematic and in the real neuron, no directional field is present in the inverted session (that is, no firing on the inner (grey) part of the toroidal manifold), as predicted above. **d**, Upper panel, schematic representation of an azimuth \times pitch cell, tuned to positive pitch (an extreme pitch neuron). Lower panel, example of an actual neuron exhibiting tuning to positive pitch, similar to the schematic. In this case, the activity of the neuron in the upright session is in fact correlated with its activity in the inverted session, as predicted above. **e**, Differences in 2D correlations between the upright and inverted session, for all the pitch-tuned neurons recorded in setup number 1 (both pure and conjunctive), computed similarly to the correlation analysis in Extended Data Fig. 7k (see Methods). This correlation was significantly larger for pitch cells that were tuned to extreme pitch (pitch $\leq -35^\circ$ or pitch $\geq +35^\circ$; 'extreme pitch', $n = 22$ cells \times sessions), compared to pitch cells tuned approximately to zero pitch (between -35° and $+35^\circ$; 'horizontal pitch', $n = 20$ cells \times sessions). Error bars, mean \pm s.e.m.; $***P < 0.001$.



Extended Data Figure 10 | Pitch cells are narrowly tuned and span a range of 360°, similar to azimuth cells. **a, b**, Example azimuth cells recorded on the arena floor in setup number 1 (panel **a**) and pitch cells recorded on the vertical ring in setup number 2 (panel **b**), showing that preferred directions of both neuronal types span the entire range of 360° (from 0° to 360° of azimuth for azimuth cells, and from -180° to +180° of pitch angles for pitch cells). Cells were sorted according to their preferred azimuth (top) or preferred pitch (bottom), highlighting the similarity of the tuning properties of azimuth cells and pitch cells. **c**, Pitch cells that were recorded on the vertical ring in two

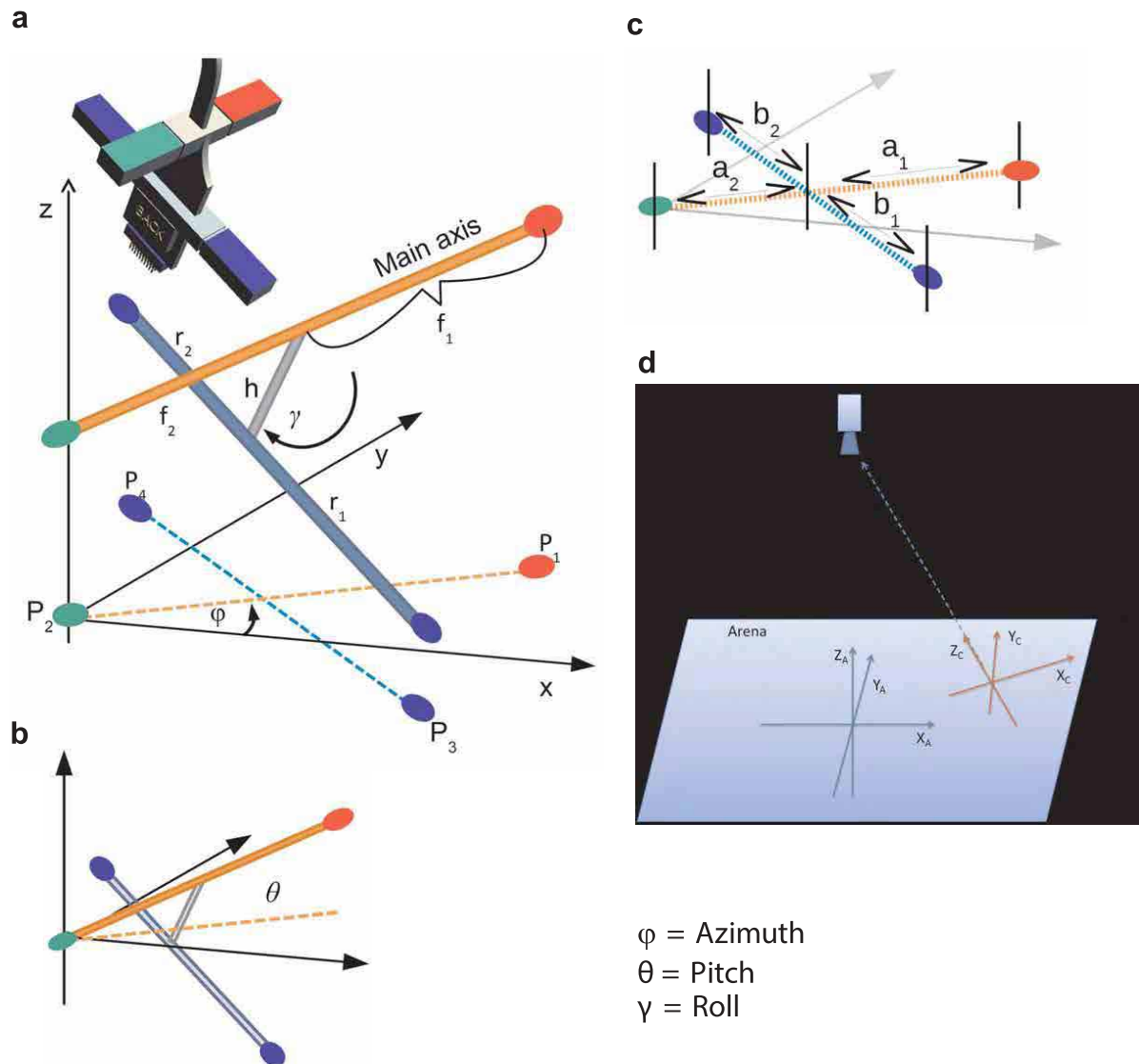
separate sessions exhibited a stable unimodal tuning. **d**, Tuning width to azimuth and to pitch. Left, population average tuning width. Right, average tuning widths for each individual animal (one bar per animal: azimuth, first 4 bars, coloured blue; pitch, last 2 bars, coloured red). The first 4 bars represent the tuning widths of azimuth cells recorded from the 4 bats in the upright-crawling experiment (setup number 1), and the last 2 bars represent the tuning widths of pitch cells recorded from the 2 bats in the vertical-ring experiment (setup number 2). Error bars, mean \pm s.e.m.



Extended Data Figure 11 | Device used for measuring the 3 Euler angles of the bat's head, based on a 3D non-coplanar arrangement of four LEDs.

a, b, Schematic illustrations of the device used for computing the head-direction angles using the top-view camera. This device was mounted on the recording headstage, and allowed measuring the Euler angles using one overhead camera (Methods). Shown are several camera-views of a schematic illustration of

the four-LED headstage, rotated in azimuth (**a**) or rotated at different combinations of pitch and roll (**b**). Central panels in both **a** and **b**: zero pitch and no roll ('flat head' position). The azimuth of the device was defined as the absolute direction of the red LED along the green-red direction. Pitch and roll were kept constant in all the plots in **a**; azimuth was kept constant in all the plots in **b**.



Extended Data Figure 12 | Definitions of intermediate angles used during the computation of the final Euler angles of the head. In order to compute the final Euler angles (in arena coordinates), we first computed intermediate Euler angles with respect to the plane of the camera, and then transformed them into arena coordinates based on the x - y position of the animal inside the arena (see Methods for the detailed computation and full definitions of these angles). **a**, **b**, Illustration of the 3D tetrahedral arrangement of 4 LEDs, including the relevant geometry and angles. **c**, Illustration of measurement as

seen by the camera. The distances relevant for the measurement of angles have been labelled a_1 , a_2 , b_1 , b_2 . **d**, Coordinate frames of the arena (' A ') and camera-view (' C '). Shown are the arena frame in blue and an example of the camera frame in red, for a particular location of the bat. Note that the alignment of the camera frame with respect to the arena frame changes as a function of the position of the LED device (position of the bat's head) within the arena; the algorithm described in the Methods section disambiguates these changes.

Correspondence

Peer review as a business transaction

I receive 300 requests a year to review research papers. Each takes 3–4 hours to complete, equivalent to roughly 1–2 days per week if I did them all. Should I win ‘reviewer of the year’ award, however, I suspect my colleagues would see my efforts as a foolish waste of time.

Reviewers are crucial to the success of prestigious and profitable journals, traditionally receiving no monetary or other recognition. As journals proliferate and scientists get ever busier, our appetite for reviewing wanes (see, for example, M. Arns *Nature* **515**, 467; 2014). One way to revive this activity would be to consider it a business transaction — with modest remuneration of, say, US\$50 per hour (see also S. Ott and D. Hebenstreit *Nature* **506**, 295; 2014).

Publishing in an open-access journal costs around \$1,000–\$2,000, so paying \$200 to a reviewer does not seem excessive. The authors and the journal could split the cost equally.

Prospective reviewers would be more inclined to do a speedy and thorough job. Retired scientists with extensive expertise and plenty of free time would be keen to participate. Editors would be spared the hunt for willing referees.

We could then use our reviewing fees to buy back some pleasure — I might go for a billiards table, a pinball machine or even a fancy treadmill.

Eleftherios P. Diamandis *Mount Sinai Hospital and University Health Network, Toronto; and University of Toronto, Canada.*
ediamandis@mthsina.on.ca

Pollution: uncouple from economy boom

The International Monetary Fund announced last year that China has overtaken the United States as the world's largest economy, accounting for 16.5% of the 2014

global gross domestic product in terms of purchasing-power parity. This achievement comes at a huge price for China.

It is well documented that China's industrial boom is causing a dangerous deterioration in air quality, as well as widespread land degradation and water and soil contamination. In 2013, 71 out of 74 Chinese cities failed to meet ambient air-quality standards (see go.nature.com/ndujl3; in Chinese).

Long-term exposure to air pollution has led to an average drop of 5 years in life expectancy for 500 million people in north China (Y. Chen *et al. Proc. Natl Acad. Sci. USA* **110**, 12936–12941; 2013). According to the World Health Organization, respiratory disease could kill more than 3 million Chinese people every year by 2030 (V. Hughes *Nature* **489**, S18–S20; 2012).

China must tackle its air pollution — to mitigate climate and as a priority for human health (see also J. Schmale *et al. Nature* **515**, 335–337; 2014).

Hong-Wei Xiao *China Agricultural University, Beijing, China.*

A. S. Mujumdar *McGill University, Quebec, Canada.*

Liming Che *Xiamen University, Xiamen, China.*
xhwcaugxy@163.com

Pollution: clear blue skies over Beijing

China's citizens coined the phrase ‘APEC blue’ to describe the Beijing sky during the Asia-Pacific Economic Cooperation summit in November last year. This transformation from the usual heavily polluted grey was a result of tough 11-day emissions-reduction measures, enforced for the benefit of visiting dignitaries.

A week or two is not enough. China must take prompt action to extend its blue-sky days for the sake of the nation's lungs and the world's climate during the economic boom (see J. Schmale *et al. Nature* **515**, 335–337; 2014) and call for concerted action by offending sectors

(see G. Yang *et al. Lancet* **381**, 1987–2015; 2013).

Zhaohui Gong *Ningbo University School of Medicine, Ningbo, China.*
zhaohui@ncrri.org.cn

Pollution: build on success in China

Before the Asia-Pacific Economic Cooperation (APEC) summit in Beijing on 5–11 November 2014, the Chinese government imposed an emergency emissions-reduction strategy to combat the city's dense smog. To deal with its environmental problems effectively, China should build on the success of this short-term intervention by establishing collaborative regional policies and compensation systems for mitigating pollution.

Production in thousands of highly polluting factories and construction sites in Beijing and its surrounding provinces was curtailed or halted for a minimum of a week before the summit. Cars with even or odd registration numbers were banned on alternate days, cutting daily traffic by some ten million vehicles.

These measures caused a sharp drop of almost 30% in the daily average concentration of particulate matter measuring up to 2.5 micrometres (PM_{2.5}) in Beijing and nearby provinces on 1–12 November. In the city itself, PM_{2.5} fell to under half the concentration for the same period in 2013 (see go.nature.com/ls8wl1; in Chinese). Blue skies appeared during the summit and remained for a week afterwards.

Our ongoing study of urban and rural development and its environmental effects in the Bohai Rim region of China will help to assess the extent to which regulating such socio-economic changes could mitigate pollution (see go.nature.com/lo4jcz; in Chinese).

Yansui Liu, Yuheng Li, Cong Chen *Beijing Normal University; and Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of*

Sciences, Beijing, China.
liyuheng@igsrr.ac.cn

International Council for Science responds

Your discussion of the challenges facing the International Council for Science paints a misleadingly negative picture (*Nature* **515**, 311; 2014).

We are a non-governmental organization representing academies and research councils from 140 countries and the science community through 31 disciplinary unions, and are a leading voice for science.

The council has initiated global research projects such as the International Polar Years. Our current flagship projects are Future Earth: Research for Global Sustainability; Urban Health and Wellbeing; and Integrated Research on Disaster Risk. The biological sciences are fully integrated into our programmes (see go.nature.com/xhzhif).

The council represents science in global organizations such as the United Nations Educational, Scientific and Cultural Organization, and is the scientific and technology lead for the UN sustainable development goals programme.

The council is dedicated to the promotion of freedom and responsibility of scientists, and champions open access to data and information through its Committee on Data for Science and Technology and new World Data System. To improve links between science and policy, the council convened a meeting of governmental science advisers (www.globalscienceadvice.org), and a global network is being established.

Details of these activities and the external review report commissioned by the council can be found at www.icsu.org.

Gordon McBean* *International Council for Science, Paris.*
gmcbean@uwo.ca

*On behalf of 15 correspondents (see go.nature.com/v591un for full list).